

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

- (i) APPLICANT: Ward, Eric  
Volrath, Sandra  
Johnson, Marie  
Potter, Sharon
- (ii) TITLE OF INVENTION: HERBICIDE TOLERANT PROTOX GENES PRODUCED BY DNA SHUFFLING
- (iii) NUMBER OF SEQUENCES: 37
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: Novartis Corporation
  - (B) STREET: 3054 Cornwallis Road
  - (C) CITY: Research Triangle Park
  - (D) STATE: NC
  - (E) COUNTRY: USA
  - (F) ZIP: 27709
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Floppy disk
  - (B) COMPUTER: IBM PC compatible
  - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
  - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
  - (A) APPLICATION NUMBER:
  - (B) FILING DATE:
  - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 09/059,164
  - (B) FILING DATE: 13-APR-1998
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 09/050,603
  - (B) FILING DATE: 30-MAR-1998
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 09/038,878
  - (B) FILING DATE: 11-MAR-1998
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 08/808,931
  - (B) FILING DATE: 28-FEB-1997
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 60/012,705
  - (B) FILING DATE: 28-FEB-1996
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 60/013,612
  - (B) FILING DATE: 28-FEB-1996
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 60/020,003
  - (B) FILING DATE: 21-JUN-1996
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 08/472,028
  - (B) FILING DATE: 06-JUN-1995

005027-5250E260

(B) FILING DATE: 16-JUN-1994

(C) REFERENCE/DOCKET NUMBER: CGC1847/CIP4

(B) TELEFAX: (919) 541-8689

[illegible]



445 450 455

AAG CCT AAT TCG ACC GAT CCA CTT AAA TTA GGA GTT AGG GTA TGG CCT 1446  
 Lys Pro Asn Ser Thr Asp Pro Leu Lys Leu Gly Val Arg Val Trp Pro  
 460 465 470

CAA GCC ATT CCT CAG TTT CTA GTT GGT CAC TTT GAT ATC CTT GAC ACG 1494  
 Gln Ala Ile Pro Gln Phe Leu Val Gly His Phe Asp Ile Leu Asp Thr  
 475 480 485

GCT AAA TCA TCT CTA ACG TCT TCG GGC TAC GAA GGG CTA TTT TTG GGT 1542  
 Ala Lys Ser Ser Leu Thr Ser Ser Gly Tyr Glu Gly Leu Phe Leu Gly  
 490 495 500

GGC AAT TAC GTC GCT GGT GTA GCC TTA GGC CGG TGT GTA GAA GGC GCA 1590  
 Gly Asn Tyr Val Ala Gly Val Ala Leu Gly Arg Cys Val Glu Gly Ala  
 505 510 515 520

TAT GAA ACC GCG ATT GAG GTC AAC AAC TTC ATG TCA CGG TAC GCT TAC 1638  
 Tyr Glu Thr Ala Ile Glu Val Asn Asn Phe Met Ser Arg Tyr Ala Tyr  
 525 530 535

AAG TAAATGTAAA ACATTAAATC TCCCAGCTTG CGTGAGTTTT ATTAAATATT 1691  
 Lys

TTGAGATATC CAAAAAAAAA AAAAAAAAAA 1719

## (2) INFORMATION FOR SEQ ID NO:2:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 537 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Glu Leu Ser Leu Leu Arg Pro Thr Thr Gln Ser Leu Leu Pro Ser  
 1 5 10 15

Phe Ser Lys Pro Asn Leu Arg Leu Asn Val Tyr Lys Pro Leu Arg Leu  
 20 25 30

Arg Cys Ser Val Ala Gly Gly Pro Thr Val Gly Ser Ser Lys Ile Glu  
 35 40 45

Gly Gly Gly Gly Thr Thr Ile Thr Thr Asp Cys Val Ile Val Gly Gly  
 50 55 60

Gly Ile Ser Gly Leu Cys Ile Ala Gln Ala Leu Ala Thr Lys His Pro  
 65 70 75 80

Asp Ala Ala Pro Asn Leu Ile Val Thr Glu Ala Lys Asp Arg Val Gly  
 85 90 95

Gly Asn Ile Ile Thr Arg Glu Glu Asn Gly Phe Leu Trp Glu Glu Gly  
 100 105 110

Pro Asn Ser Phe Gln Pro Ser Asp Pro Met Leu Thr Met Val Val Asp  
 115 120 125

Ser Gly Leu Lys Asp Asp Leu Val Leu Gly Asp Pro Thr Ala Pro Arg

130 135 140  
 Phe Val Leu Trp Asn Gly Lys Leu Arg Pro Val Pro Ser Lys Leu Thr  
 145 150 155 160  
 Asp Leu Pro Phe Phe Asp Leu Met Ser Ile Gly Gly Lys Ile Arg Ala  
 165 170 175  
 Gly Phe Gly Ala Leu Gly Ile Arg Pro Ser Pro Pro Gly Arg Glu Glu  
 180 185 190  
 Ser Val Glu Glu Phe Val Arg Arg Asn Leu Gly Asp Glu Val Phe Glu  
 195 200 205  
 Arg Leu Ile Glu Pro Phe Cys Ser Gly Val Tyr Ala Gly Asp Pro Ser  
 210 215 220  
 Lys Leu Ser Met Lys Ala Ala Phe Gly Lys Val Trp Lys Leu Glu Gln  
 225 230 235 240  
 Asn Gly Gly Ser Ile Ile Gly Gly Thr Phe Lys Ala Ile Gln Glu Arg  
 245 250 255  
 Lys Asn Ala Pro Lys Ala Glu Arg Asp Pro Arg Leu Pro Lys Pro Gln  
 260 265 270  
 Gly Gln Thr Val Gly Ser Phe Arg Lys Gly Leu Arg Met Leu Pro Glu  
 275 280 285  
 Ala Ile Ser Ala Arg Leu Gly Ser Lys Val Lys Leu Ser Trp Lys Leu  
 290 295 300  
 Ser Gly Ile Thr Lys Leu Glu Ser Gly Gly Tyr Asn Leu Thr Tyr Glu  
 305 310 315 320  
 Thr Pro Asp Gly Leu Val Ser Val Gln Ser Lys Ser Val Val Met Thr  
 325 330 335  
 Val Pro Ser His Val Ala Ser Gly Leu Leu Arg Pro Leu Ser Glu Ser  
 340 345 350  
 Ala Ala Asn Ala Leu Ser Lys Leu Tyr Tyr Pro Pro Val Ala Ala Val  
 355 360 365  
 Ser Ile Ser Tyr Pro Lys Glu Ala Ile Arg Thr Glu Cys Leu Ile Asp  
 370 375 380  
 Gly Glu Leu Lys Gly Phe Gly Gln Leu His Pro Arg Thr Gln Gly Val  
 385 390 395 400  
 Glu Thr Leu Gly Thr Ile Tyr Ser Ser Ser Leu Phe Pro Asn Arg Ala  
 405 410 415  
 Pro Pro Gly Arg Ile Leu Leu Leu Asn Tyr Ile Gly Gly Ser Thr Asn  
 420 425 430  
 Thr Gly Ile Leu Ser Lys Ser Glu Gly Glu Leu Val Glu Ala Val Asp  
 435 440 445  
 Arg Asp Leu Arg Lys Met Leu Ile Lys Pro Asn Ser Thr Asp Pro Leu  
 450 455 460  
 Lys Leu Gly Val Arg Val Trp Pro Gln Ala Ile Pro Gln Phe Leu Val  
 465 470 475 480

005021-5250E450

(2) INFORMATION FOR SEO ID NO:3:

(A) LENGTH: 1738 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(vii) IMMEDIATE SOURCE:

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 70..1596

(D) OTHER INFORMATION: /product= "Arabidopsis protox-2"

TTTTTTACTT	ATTTCCGTC	CTGCTTTCTGA	CTGGTCAGAG	ATTTTGACTC	TGAATTGTTG	60
CAGATAGCA	ATG GCG TCT GGA GCA GTA GCA GAT CAT CAA ATT GAA GCG	108				
	Met Ala Ser Gly Ala Val Ala Asp His Gln Ile Glu Ala					
	1 5 10					
GTT TCA GGA AAA AGA GTC GCA GTC GTA GGT GCA GGT GTA AGT GGA CTT	156					
Val Ser Gly Lys Arg Val Ala Val Val Gly Ala Gly Val Ser Gly Leu						
	15 20 25					
GCG GCG GCT TAC AAG TTG AAA TCG AGG GGT TTG AAT GTG ACT GTG TTT	204					
Ala Ala Ala Tyr Lys Leu Lys Ser Arg Gly Leu Asn Val Thr Val Phe						
	30 35 40 45					
GAA GCT GAT GGA AGA GTA GGT GGG AAG TTG AGA AGT GTT ATG CAA AAT	252					
Glu Ala Asp Gly Arg Val Gly Gly Lys Leu Arg Ser Val Met Gln Asn						
	50 55 60					
GGT TTG ATT TGG GAT GAA GGA GCA AAC ACC ATG ACT GAG GCT GAG CCA	300					
Gly Leu Ile Trp Asp Glu Gly Ala Asn Thr Met Thr Glu Ala Glu Pro						
	65 70 75					
GAA GTT GGG AGT TTA CTT GAT GAT CTT GGG CTT CGT GAG AAA CAA CAA	348					
Glu Val Gly Ser Leu Leu Asp Asp Leu Gly Leu Arg Glu Lys Gln Gln						



(2) INFORMATION FOR SEO ID NO:4:

(A) LENGTH: 508 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEO ID NO:4:

135



Tyr Lys Leu Lys Ser Arg Gly Leu Asn Val Thr Val Phe Glu Ala Asp  
           35                          40                          45  
 Gly Arg Val Gly Gly Lys Leu Arg Ser Val Met Gln Asn Gly Leu Ile  
           50                          55                          60  
 Trp Asp Glu Gly Ala Asn Thr Met Thr Glu Ala Glu Pro Glu Val Gly  
           65                          70                          75                          80  
 Ser Leu Leu Asp Asp Leu Gly Leu Arg Glu Lys Gln Gln Phe Pro Ile  
                           85                          90                          95  
 Ser Gln Lys Lys Arg Tyr Ile Val Arg Asn Gly Val Pro Val Met Leu  
                           100                          105                          110  
 Pro Thr Asn Pro Ile Glu Leu Val Thr Ser Ser Val Leu Ser Thr Gln  
                           115                          120                          125  
 Ser Lys Phe Gln Ile Leu Leu Glu Pro Phe Leu Trp Lys Lys Lys Ser  
           130                          135                          140  
 Ser Lys Val Ser Asp Ala Ser Ala Glu Glu Ser Val Ser Glu Phe Phe  
           145                          150                          155                          160  
 Gln Arg His Phe Gly Gln Glu Val Val Asp Tyr Leu Ile Asp Pro Phe  
                           165                          170                          175  
 Val Gly Gly Thr Ser Ala Ala Asp Pro Asp Ser Leu Ser Met Lys His  
                           180                          185                          190  
 Ser Phe Pro Asp Leu Trp Asn Val Glu Lys Ser Phe Gly Ser Ile Ile  
                           195                          200                          205  
 Val Gly Ala Ile Arg Thr Lys Phe Ala Ala Lys Gly Gly Lys Ser Arg  
           210                          215                          220  
 Asp Thr Lys Ser Ser Pro Gly Thr Lys Lys Gly Ser Arg Gly Ser Phe  
           225                          230                          235                          240  
 Ser Phe Lys Gly Gly Met Gln Ile Leu Pro Asp Thr Leu Cys Lys Ser  
                           245                          250                          255  
 Leu Ser His Asp Glu Ile Asn Leu Asp Ser Lys Val Leu Ser Leu Ser  
                           260                          265                          270  
 Tyr Asn Ser Gly Ser Arg Gln Glu Asn Trp Ser Leu Ser Cys Val Ser  
           275                          280                          285  
 His Asn Glu Thr Gln Arg Gln Asn Pro His Tyr Asp Ala Val Ile Met  
           290                          295                          300  
 Thr Ala Pro Leu Cys Asn Val Lys Glu Met Lys Val Met Lys Gly Gly  
           305                          310                          315                          320  
 Gln Pro Phe Gln Leu Asn Phe Leu Pro Glu Ile Asn Tyr Met Pro Leu  
                           325                          330                          335  
 Ser Val Leu Ile Thr Thr Phe Thr Lys Glu Lys Val Lys Arg Pro Leu  
                           340                          345                          350  
 Glu Gly Phe Gly Val Leu Ile Pro Ser Lys Glu Gln Lys His Gly Phe  
           355                          360                          365  
 Lys Thr Leu Gly Thr Leu Phe Ser Ser Met Met Phe Pro Asp Arg Ser  
           370                          375                          380

005021-22506260

Pro Ser Asp Val His Leu Tyr Thr Thr Phe Ile Gly Gly Ser Arg Asn  
385 390 395 400

Gln Glu Leu Ala Lys Ala Ser Thr Asp Glu Leu Lys Gln Val Val Thr  
405 410 415

Ser Asp Leu Gln Arg Leu Leu Gly Val Glu Gly Glu Pro Val Ser Val  
420 425 430

Asn His Tyr Tyr Trp Arg Lys Ala Phe Pro Leu Tyr Asp Ser Ser Tyr  
435 440 445

Asp Ser Val Met Glu Ala Ile Asp Lys Met Glu Asn Asp Leu Pro Gly  
450 455 460

Phe Phe Tyr Ala Gly Asn His Arg Gly Gly Leu Ser Val Gly Lys Ser  
465 470 475 480

Ile Ala Ser Gly Cys Lys Ala Ala Asp Leu Val Ile Ser Tyr Leu Glu  
485 490 495

Ser Cys Ser Asn Asp Lys Lys Pro Asn Asp Ser Leu  
500 505

## (2) INFORMATION FOR SEQ ID NO:5:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1691 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## (iii) HYPOTHETICAL: NO

## (iv) ANTI-SENSE: NO

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Zea mays (maize)

## (vii) IMMEDIATE SOURCE:

- (B) CLONE: pWDC-4 (NRRL B-21260)

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..1443
- (D) OTHER INFORMATION: /product= "Maize protox-1

cDNA "

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GCG GAC TGC GTC GTG GTG GGC GGA GGC ATC AGT GGC CTC TGC ACC GCG	48
Ala Asp Cys Val Val Val Gly Gly Gly Ile Ser Gly Leu Cys Thr Ala	
1 5 10 15	
CAG GCG CTG GCC ACG CGG CAC GGC GTC GGG GAC GTG CTT GTC ACG GAG	96
Gln Ala Leu Ala Thr Arg His Gly Val Gly Asp Val Leu Val Thr Glu	
20 25 30	
GCC CGC GCC CGC CCC GGC GGC AAC ATT ACC ACC GTC GAG CGC CCC GAG	144
Ala Arg Ala Arg Pro Gly Gly Asn Ile Thr Thr Val Glu Arg Pro Glu	

35	40	45	
GAA GGG TAC CTC TGG GAG GAG GGT CCC AAC AGC TTC CAG CCC TCC GAC Glu Gly Tyr Leu Trp Glu Glu Gly Pro Asn Ser Phe Gln Pro Ser Asp 50 55 60			192
CCC GTT CTC ACC ATG GCC GTG GAC AGC GGA CTG AAG GAT GAC TTG GTT Pro Val Leu Thr Met Ala Val Asp Ser Gly Leu Lys Asp Asp Leu Val 65 70 75 80			240
TTT GGG GAC CCA AAC GCG CCG CGT TTC GTG CTG TGG GAG GGG AAG CTG Phe Gly Asp Pro Asn Ala Pro Arg Phe Val Leu Trp Glu Gly Lys Leu 85 90 95			288
AGG CCC GTG CCA TCC AAG CCC GCC GAC CTC CCG TTC TTC GAT CTC ATG Arg Pro Val Pro Ser Lys Pro Ala Asp Leu Pro Phe Phe Asp Leu Met 100 105 110			336
AGC ATC CCA GGG AAG CTC AGG GCC GGT CTA GGC GCG CTT GGC ATC CGC Ser Ile Pro Gly Lys Leu Arg Ala Gly Leu Gly Ala Leu Gly Ile Arg 115 120 125			384
CCG CCT CCT CCA GGC CGC GAA GAG TCA GTG GAG GAG TTC GTG CGC CGC Pro Pro Pro Pro Gly Arg Glu Glu Ser Val Glu Glu Phe Val Arg Arg 130 135 140			432
AAC CTC GGT GCT GAG GTC TTT GAG CGC CTC ATT GAG CCT TTC TGC TCA Asn Leu Gly Ala Glu Val Phe Glu Arg Leu Ile Glu Pro Phe Cys Ser 145 150 155 160			480
GGT GTC TAT GCT GGT GAT CCT TCT AAG CTC AGC ATG AAG GCT GCA TTT Gly Val Tyr Ala Gly Asp Pro Ser Lys Leu Ser Met Lys Ala Ala Phe 165 170 175			528
GGG AAG GTT TGG CGG TTG GAA GAA ACT GGA GGT AGT ATT ATT GGT GGA Gly Lys Val Trp Arg Leu Glu Glu Thr Gly Gly Ser Ile Ile Gly Gly 180 185 190			576
ACC ATC AAG ACA ATT CAG GAG AGG AGC AAG AAT CCA AAA CCA CCG AGG Thr Ile Lys Thr Ile Gln Glu Arg Ser Lys Asn Pro Lys Pro Pro Arg 195 200 205			624
GAT GCC CGC CTT CCG AAG CCA AAA GGG CAG ACA GTT GCA TCT TTC AGG Asp Ala Arg Leu Pro Lys Pro Lys Gly Gln Thr Val Ala Ser Phe Arg 210 215 220			672
AAG GGT CTT GCC ATG CTT CCA AAT GCC ATT ACA TCC AGC TTG GGT AGT Lys Gly Leu Ala Met Leu Pro Asn Ala Ile Thr Ser Ser Leu Gly Ser 225 230 235 240			720
AAA GTC AAA CTA TCA TGG AAA CTC ACG AGC ATT ACA AAA TCA GAT GAC Lys Val Lys Leu Ser Trp Lys Leu Thr Ser Ile Thr Lys Ser Asp Asp 245 250 255			768
AAG GGA TAT GTT TTG GAG TAT GAA ACG CCA GAA GGG GTT GTT TCG GTG Lys Gly Tyr Val Leu Glu Tyr Glu Thr Pro Glu Gly Val Val Ser Val 260 265 270			816
CAG GCT AAA AGT GTT ATC ATG ACT ATT CCA TCA TAT GTT GCT AGC AAC Gln Ala Lys Ser Val Ile Met Thr Ile Pro Ser Tyr Val Ala Ser Asn 275 280 285			864
ATT TTG CGT CCA CTT TCA AGC GAT GCT GCA GAT GCT CTA TCA AGA TTC Ile Leu Arg Pro Leu Ser Ser Asp Ala Ala Asp Ala Leu Ser Arg Phe 290 295 300			912

005021 52502460

TAT Tyr 305	TAT Tyr 310	CCA Pro 315	CCG Pro 320	GTT Val 325	GCT Ala 330	GCT Ala 335	GTA Val 340	ACT Thr 345	GTT Val 350	TCG Ser 355	TAT Tyr 360	CCA Pro 365	AAG Lys 370	GAA Glu 375	GCA Ala 380	960	
ATT Ile 325	AGA Arg 330	AAA Lys 335	GAA Glu 340	TGC Cys 345	TTA Leu 350	ATT Ile 355	GAT Asp 360	GGG Gly 365	GAA Glu 370	CTC Leu 375	CAG Gln 380	GGC Gly 385	TTT Phe 390	GGC Gly 395	CAG Gln 400	1008	
TTG Leu 340	CAT His 345	CCA Pro 350	CGT Arg 355	AGT Ser 360	CAA Gln 365	GGA Gly 370	GTT Val 375	GAG Glu 380	ACA Thr 385	TTA Leu 390	GGA Gly 395	ACA Val 400	ATA Ile 405	TAC Tyr 410	AGT Ser 415	1056	
TCC Ser 355	TCA Ser 360	CTC Leu 365	TTT Phe 370	CCA Pro 375	AAT Asn 380	CGT Arg 385	GCT Ala 390	CCT Pro 395	GAC Asp 400	GGT Gly 405	AGG Arg 410	GTG Val 415	TTA Leu 420	CTT Leu 425	CTA Leu 430	1104	
AAC Asn 370	TAC Tyr 375	ATA Ile 380	GGA Gly 385	GGT Gly 390	GCT Ala 395	ACA Thr 400	AAC Asn 405	ACA Thr 410	GGA Gly 415	ATT Ile 420	GTT Val 425	TCC Ser 430	AAG Lys 435	ACT Thr 440	GAA Glu 445	1152	
AGT Ser 385	GAG Glu 390	CTG Leu 395	GTC Val 400	GAA Glu 405	GCA Ala 410	GTT Val 415	GAC Asp 420	CGT Arg 425	GAC Asp 430	CTC Leu 435	CGA Arg 440	AAA Lys 445	ATG Met 450	CTT Leu 455	ATA Ile 460	1200	
AAT Asn 405	TCT Ser 410	ACA Thr 415	GCA Ala 420	GTG Val 425	GAC Asp 430	CCT Pro 435	TTA Leu 440	GTC Val 445	CTT Leu 450	GGT Gly 455	GTT Val 460	CGA Arg 465	GTT Val 470	TGG Trp 475	CCA Pro 480	1248	
CAA Gln 420	GCC Ala 425	ATA Ile 430	CCT Pro 435	CAG Gln 440	TTC Phe 445	CTG Leu 450	GTA Val 455	GGA Gly 460	CAT His 465	CTT Leu 470	GAT Asp 475	CTT Leu 480	CTG Leu 485	GAA Glu 490	GCC Ala 495	1296	
GCA Ala 435	AAA Lys 440	GCT Ala 445	GCC Ala 450	CTG Leu 455	GAC Asp 460	CGA Arg 465	GGT Gly 470	GGC Gly 475	TAC Tyr 480	GAT Asp 485	GGG Gly 490	CTG Leu 495	TTC Phe 500	CTA Leu 505	GGA Gly 510	1344	
GGG Gly 450	AAC Asn 455	TAT Tyr 460	GTT Val 465	GCA Ala 470	GGA Gly 475	GTT Val 480	GCC Ala 485	CTG Leu 490	GGC Gly 495	AGA Arg 500	TGC Cys 505	GTT Val 510	GAG Glu 515	GGC Gly 520	GCG Ala 525	1392	
TAT Tyr 465	GAA Glu 470	AGT Ser 475	GCC Ala 480	TCG Ser 485	CAA Gln 490	ATA Ile 495	TCT Ser 500	GAC Asp 505	TTC Phe 510	TTG Leu 515	ACC Thr 520	AAG Lys 525	TAT Tyr 530	GCC Ala 535	TAC Tyr 540	1440	
AAG Lys	TGATGAAAGA	AGTGGAGCGC	TACTTGTTAA	TCGTTTATGT	TGCATAGATG												1493
AGGTGCCTCC	GGGGAAAAAA	AAGCTTGAAT	AGTATTTTTT	ATTCTTATTT	TGTAAATTGC												1553
ATTTCTGTTC	TTTTTCTAT	CAGTAATTAG	TTATATTTTA	GTTCTGTAGG	AGATTGTTCT												1613
GTTCACTGCC	CTTCAAAGA	AATTTTATTT	TTCATTCTTT	TATGAGAGCT	GTGCTACTTA												1673
AAAAAAAAAA	AAAAAAAA																1691

## (2) INFORMATION FOR SEQ ID NO:6:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 481 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Ala Asp Cys Val Val Val Gly Gly Gly Ile Ser Gly Leu Cys Thr Ala  
 1 5 10 15  
 Gln Ala Leu Ala Thr Arg His Gly Val Gly Asp Val Leu Val Thr Glu  
 20 25 30  
 Ala Arg Ala Arg Pro Gly Gly Asn Ile Thr Thr Val Glu Arg Pro Glu  
 35 40 45  
 Glu Gly Tyr Leu Trp Glu Glu Gly Pro Asn Ser Phe Gln Pro Ser Asp  
 50 55 60  
 Pro Val Leu Thr Met Ala Val Asp Ser Gly Leu Lys Asp Asp Leu Val  
 65 70 75 80  
 Phe Gly Asp Pro Asn Ala Pro Arg Phe Val Leu Trp Glu Gly Lys Leu  
 85 90 95  
 Arg Pro Val Pro Ser Lys Pro Ala Asp Leu Pro Phe Phe Asp Leu Met  
 100 105 110  
 Ser Ile Pro Gly Lys Leu Arg Ala Gly Leu Gly Ala Leu Gly Ile Arg  
 115 120 125  
 Pro Pro Pro Pro Gly Arg Glu Glu Ser Val Glu Glu Phe Val Arg Arg  
 130 135 140  
 Asn Leu Gly Ala Glu Val Phe Glu Arg Leu Ile Glu Pro Phe Cys Ser  
 145 150 155 160  
 Gly Val Tyr Ala Gly Asp Pro Ser Lys Leu Ser Met Lys Ala Ala Phe  
 165 170 175  
 Gly Lys Val Trp Arg Leu Glu Glu Thr Gly Gly Ser Ile Ile Gly Gly  
 180 185 190  
 Thr Ile Lys Thr Ile Gln Glu Arg Ser Lys Asn Pro Lys Pro Pro Arg  
 195 200 205  
 Asp Ala Arg Leu Pro Lys Pro Lys Gly Gln Thr Val Ala Ser Phe Arg  
 210 215 220  
 Lys Gly Leu Ala Met Leu Pro Asn Ala Ile Thr Ser Ser Leu Gly Ser  
 225 230 235 240  
 Lys Val Lys Leu Ser Trp Lys Leu Thr Ser Ile Thr Lys Ser Asp Asp  
 245 250 255  
 Lys Gly Tyr Val Leu Glu Tyr Glu Thr Pro Glu Gly Val Val Ser Val  
 260 265 270  
 Gln Ala Lys Ser Val Ile Met Thr Ile Pro Ser Tyr Val Ala Ser Asn  
 275 280 285  
 Ile Leu Arg Pro Leu Ser Ser Asp Ala Ala Asp Ala Leu Ser Arg Phe  
 290 295 300  
 Tyr Tyr Pro Pro Val Ala Ala Val Thr Val Ser Tyr Pro Lys Glu Ala  
 305 310 315 320  
 Ile Arg Lys Glu Cys Leu Ile Asp Gly Glu Leu Gln Gly Phe Gly Gln

(2) INFORMATION FOR SEQ ID NO:7:

(A) LENGTH: 2061 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *Zea mays* (maize)

(vii) IMMEDIATE SOURCE:

(B) CLONE: pWDC-3 (NRRL B-21259)

(ix) **FEATURE:**

(A) NAME/KEY: CDS

(B) LOCATION: 64..1698

(D) OTHER INFORMATION: /product= "Maize protox-2"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

P:\ABRU\1800A\1847CIP4\1847CIP4.DOC

142





Asp Leu Ala Ile Ser Tyr Leu Glu Ser His Thr Lys His Asn Asn Ser  
530 535 540

CAT TGAAAGTGTC TGACCTATCC TCTAGCAGTT GTCGACAAAT TTCTCCAGTT 1745  
His  
545

CATGTACAGT AGAAACCGAT GCGTTGCAGT TTCAGAACAT CTTCACTTCT TCAGATATTA 1805  
ACCCTTCGTT GAACATCCAC CAGAAAGGTA GTCACATGTG TAAGTGGGAA AATGAGGTTA 1865  
AAACTATTA TGGCGGCCGA AATGTTTCCTT TTTGTTTTCC TCACAAGTGG CCTACGACAC 1925  
TTGATGTTGG AAATACATTT AAATTTGTTG AATTGTTTGA GAACACATGC GTGACGTGTA 1985  
ATATTTGCCT ATTGTGATTT TAGCAGTAGT CTTGGCCAGA TTATGCTTTA CGCCTTTAAA 2045  
AAAAAAAAA AAAAAA 2061

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 544 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met Leu Ala Leu Thr Ala Ser Ala Ser Ser Ala Ser Ser His Pro Tyr  
1 5 10 15  
Arg His Ala Ser Ala His Thr Arg Arg Pro Arg Leu Arg Ala Val Leu  
20 25 30  
Ala Met Ala Gly Ser Asp Asp Pro Arg Ala Ala Pro Ala Arg Ser Val  
35 40 45  
Ala Val Val Gly Ala Gly Val Ser Gly Leu Ala Ala Ala Tyr Arg Leu  
50 55 60  
Arg Gln Ser Gly Val Asn Val Thr Val Phe Glu Ala Ala Asp Arg Ala  
65 70 75 80  
Gly Gly Lys Ile Arg Thr Asn Ser Glu Gly Gly Phe Val Trp Asp Glu  
85 90 95  
Gly Ala Asn Thr Met Thr Glu Gly Glu Trp Glu Ala Ser Arg Leu Ile  
100 105 110  
Asp Asp Leu Gly Leu Gln Asp Lys Gln Gln Tyr Pro Asn Ser Gln His  
115 120 125  
Lys Arg Tyr Ile Val Lys Asp Gly Ala Pro Ala Leu Ile Pro Ser Asp  
130 135 140  
Pro Ile Ser Leu Met Lys Ser Ser Val Leu Ser Thr Lys Ser Lys Ile  
145 150 155 160  
Ala Leu Phe Phe Glu Pro Phe Leu Tyr Lys Lys Ala Asn Thr Arg Asn  
165 170 175  
Ser Gly Lys Val Ser Glu Glu His Leu Ser Glu Ser Val Gly Ser Phe

145

Leu Ala Ile Ser Tyr Leu Glu Ser His Thr Lys His Asn Asn Ser His  
 530 535 540

## (2) INFORMATION FOR SEQ ID NO:9:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1811 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

## (vi) ORIGINAL SOURCE:

(A) ORGANISM: *Triticum aestivum* (wheat)

## (vii) IMMEDIATE SOURCE:

(B) CLONE: pWDC-13 (NRRL B-21545)

## (ix) FEATURE:

- (A) NAME/KEY: CDS  
 (B) LOCATION: 3..1589  
 (D) OTHER INFORMATION: /product= "wheat protox-1"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GC GCA ACA ATG GCC ACC GCC ACC GTC GCG GCC GCG TCG CCG CTC CGC	47
Ala Thr Met Ala Thr Ala Thr Val Ala Ala Ala Ser Pro Leu Arg	
1 5 10 15	
GGC AGG GTC ACC GGG CGC CCA CAC CGC GTC CGC CCG CGT TGC GCT ACC	95
Gly Arg Val Thr Gly Arg Pro His Arg Val Arg Pro Arg Cys Ala Thr	
20 25 30	
GCG AGC AGC GCG ACC GAG ACT CCG GCG GCG CCC GGC GTG CCG CTG TCC	143
Ala Ser Ser Ala Thr Glu Thr Pro Ala Ala Pro Gly Val Arg Leu Ser	
35 40 45	
GCG GAA TGC GTC ATT GTG GGC GCC GGC ATC AGC GGC CTC TGC ACC GCG	191
Ala Glu Cys Val Ile Val Gly Ala Gly Ile Ser Gly Leu Cys Thr Ala	
50 55 60	
CAG GCG CTG GCC ACC CGA TAC GGC GTC AGC GAC CTG CTC GTC ACG GAG	239
Gln Ala Leu Ala Thr Arg Tyr Gly Val Ser Asp Leu Leu Val Thr Glu	
65 70 75	
GCC CGC GAC CGC CCG GGC GGC AAC ATC ACC ACC GTC GAG CGT CCC GAC	287
Ala Arg Asp Arg Pro Gly Gly Asn Ile Thr Thr Val Glu Arg Pro Asp	
80 85 90 95	
GAG GGG TAC CTG TGG GAG GAG GGA CCC AAC AGC TTC CAG CCC TCC GAC	335
Glu Gly Tyr Leu Trp Glu Glu Gly Pro Asn Ser Phe Gln Pro Ser Asp	
100 105 110	
CCG GTC CTC ACC ATG GCC GTG GAC AGC GGG CTC AAG GAT GAC TTG GTG	383
Pro Val Leu Thr Met Ala Val Asp Ser Gly Leu Lys Asp Asp Leu Val	
115 120 125	
TTC GGG GAC CCC AAC GCG CCC CGG TTC GTG CTG TGG GAG GGG AAG CTG	431
Phe Gly Asp Pro Asn Ala Pro Arg Phe Val Leu Trp Glu Gly Lys Leu	



TCT TCT CTC TTT CCT AAT CGT GCT CCT GCT GGA AGA GTG TTA CTT CTG 1247  
 Ser Ser Leu Phe Pro Asn Arg Ala Pro Ala Gly Arg Val Leu Leu Leu  
 400 405 410 415  
 AAC TAT ATC GGG GGT TCT ACA AAT ACA GGG ATC GTC TCC AAG ACT GAG 1295  
 Asn Tyr Ile Gly Gly Ser Thr Asn Thr Gly Ile Val Ser Lys Thr Glu  
 420 425 430  
 AGT GAC TTA GTA GGA GCC GTT GAC CGT GAC CTC AGA AAA ATG TTG ATA 1343  
 Ser Asp Leu Val Gly Ala Val Asp Arg Asp Leu Arg Lys Met Leu Ile  
 435 440 445  
 AAC CCT AGA GCA GCA GAC CCT TTA GCA TTA GGG GTT CGA GTG TGG CCA 1391  
 Asn Pro Arg Ala Ala Asp Pro Leu Ala Leu Gly Val Arg Val Trp Pro  
 450 455 460  
 CAA GCA ATA CCA CAG TTT TTG ATT GGG CAC CTT GAT CGC CTT GCT GCT 1439  
 Gln Ala Ile Pro Gln Phe Leu Ile Gly His Leu Asp Arg Leu Ala Ala  
 465 470 475  
 GCA AAA TCT GCA CTG GGC CAA GGC GGC TAC GAC GGG TTG TTC CTA GGA 1487  
 Ala Lys Ser Ala Leu Gly Gln Gly Gly Tyr Asp Gly Leu Phe Leu Gly  
 480 485 490 495  
 GGA AAC TAC GTC GCA GGA GTT GCC TTG GGC CGA TGC ATC GAG GGT GCG 1535  
 Gly Asn Tyr Val Ala Gly Val Ala Leu Gly Arg Cys Ile Glu Gly Ala  
 500 505 510  
 TAC GAG AGT GCC TCA CAA GTA TCT GAC TTC TTG ACC AAG TAT GCC TAC 1583  
 Tyr Glu Ser Ala Ser Gln Val Ser Asp Phe Leu Thr Lys Tyr Ala Tyr  
 515 520 525  
 AAG TGA TGGAAGTAGT GCATCTCTTC ATTTTGTTC ATATACGAGG TGAGGCTAGG 1639  
 Lys  
 ATCGGTAAAA CATCATGAGA TTCTGTAGTG TTTCTTTAAT TGAAAAAACA AATTTTAGTG 1699  
 ATGCAATATG TGCTCTTTCC TGTAGTTCGA GCATGTACAT CGGTATGGGA TAAAGTAGAA 1759  
 TAAGCTATTC TGCAAAGCA GTGATTTTTT TTGAAAAAAA AAAAAAAAAA AA 1811

## (2) INFORMATION FOR SEQ ID NO:10:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 528 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Ala Thr Met Ala Thr Ala Thr Val Ala Ala Ala Ser Pro Leu Arg Gly  
 1 5 10 15  
 Arg Val Thr Gly Arg Pro His Arg Val Arg Pro Arg Cys Ala Thr Ala  
 20 25 30  
 Ser Ser Ala Thr Glu Thr Pro Ala Ala Pro Gly Val Arg Leu Ser Ala  
 35 40 45  
 Glu Cys Val Ile Val Gly Ala Gly Ile Ser Gly Leu Cys Thr Ala Gln

50                      55                      60  
 Ala Leu Ala Thr Arg Tyr Gly Val Ser Asp Leu Leu Val Thr Glu Ala  
 65                      70                      75                      80  
 Arg Asp Arg Pro Gly Gly Asn Ile Thr Thr Val Glu Arg Pro Asp Glu  
 85                      90                      95  
 Gly Tyr Leu Trp Glu Glu Gly Pro Asn Ser Phe Gln Pro Ser Asp Pro  
 100                      105                      110  
 Val Leu Thr Met Ala Val Asp Ser Gly Leu Lys Asp Asp Leu Val Phe  
 115                      120                      125  
 Gly Asp Pro Asn Ala Pro Arg Phe Val Leu Trp Glu Gly Lys Leu Arg  
 130                      135                      140  
 Pro Val Pro Ser Lys Pro Gly Asp Leu Pro Phe Phe Ser Leu Met Ser  
 145                      150                      155                      160  
 Ile Pro Gly Lys Leu Arg Ala Gly Leu Gly Ala Leu Gly Ile Arg Pro  
 165                      170                      175  
 Pro Pro Pro Gly Arg Glu Glu Ser Val Glu Glu Phe Val Arg Arg Asn  
 180                      185                      190  
 Leu Gly Ala Glu Val Phe Glu Arg Leu Ile Glu Pro Phe Cys Ser Gly  
 195                      200                      205  
 Val Tyr Ala Gly Asp Pro Ser Lys Leu Ser Met Lys Ala Ala Phe Gly  
 210                      215                      220  
 Lys Val Trp Arg Leu Glu Glu Ile Gly Gly Ser Ile Ile Gly Gly Thr  
 225                      230                      235                      240  
 Ile Lys Ala Ile Gln Asp Lys Gly Lys Asn Pro Lys Pro Pro Arg Asp  
 245                      250                      255  
 Pro Arg Leu Pro Ala Pro Lys Gly Gln Thr Val Ala Ser Phe Arg Lys  
 260                      265                      270  
 Gly Leu Ala Met Leu Pro Asn Ala Ile Ala Ser Arg Leu Gly Ser Lys  
 275                      280                      285  
 Val Lys Leu Ser Trp Lys Leu Thr Ser Ile Thr Lys Ala Asp Asn Gln  
 290                      295                      300  
 Gly Tyr Val Leu Gly Tyr Glu Thr Pro Glu Gly Leu Val Ser Val Gln  
 305                      310                      315                      320  
 Ala Lys Ser Val Ile Met Thr Ile Pro Ser Tyr Val Ala Ser Asp Ile  
 325                      330                      335  
 Leu Arg Pro Leu Ser Ile Asp Ala Ala Asp Ala Leu Ser Lys Phe Tyr  
 340                      345                      350  
 Tyr Pro Pro Val Ala Ala Val Thr Val Ser Tyr Pro Lys Glu Ala Ile  
 355                      360                      365  
 Arg Lys Glu Cys Leu Ile Asp Gly Glu Leu Gln Gly Phe Gly Gln Leu  
 370                      375                      380  
 His Pro Arg Ser Gln Gly Val Glu Thr Leu Gly Thr Ile Tyr Ser Ser  
 385                      390                      395                      400

005021 " 22502260

Ser Leu Phe Pro Asn Arg Ala Pro Ala Gly Arg Val Leu Leu Leu Asn  
405 410 415  
Tyr Ile Gly Gly Ser Thr Asn Thr Gly Ile Val Ser Lys Thr Glu Ser  
420 425 430  
Asp Leu Val Gly Ala Val Asp Arg Asp Leu Arg Lys Met Leu Ile Asn  
435 440 445  
Pro Arg Ala Ala Asp Pro Leu Ala Leu Gly Val Arg Val Trp Pro Gln  
450 455 460  
Ala Ile Pro Gln Phe Leu Ile Gly His Leu Asp Arg Leu Ala Ala Ala  
465 470 475 480  
Lys Ser Ala Leu Gly Gln Gly Gly Tyr Asp Gly Leu Phe Leu Gly Gly  
485 490 495  
Asn Tyr Val Ala Gly Val Ala Leu Gly Arg Cys Ile Glu Gly Ala Tyr  
500 505 510  
Glu Ser Ala Ser Gln Val Ser Asp Phe Leu Thr Lys Tyr Ala Tyr Lys  
515 520 525

## (2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1847 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: soybean
- (vii) IMMEDIATE SOURCE:
  - (B) CLONE: pWDC-12 (NRRL B-21516)
- (ix) FEATURE:
  - (A) NAME/KEY: CDS
  - (B) LOCATION: 55..1683
  - (D) OTHER INFORMATION: /product= "soybean protox-1"
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

CTTTAGCACA GTGTTGAAGA TAACGAACGA ATAGTGCCAT TACTGTAACC AACC ATG	57
Met	
1	
GTT TCC GTC TTC AAC GAG ATC CTA TTC CCG CCG AAC CAA ACC CTT CTT	105
Val Ser Val Phe Asn Glu Ile Leu Phe Pro Pro Asn Gln Thr Leu Leu	
5 10 15	
CGC CCC TCC CTC CAT TCC CCA ACC TCT TTC TTC ACC TCT CCC ACT CGA	153
Arg Pro Ser Leu His Ser Pro Thr Ser Phe Phe Thr Ser Pro Thr Arg	
20 25 30	
AAA TTC CCT CGC TCT CGC CCT AAC CCT ATT CTA CGC TGC TCC ATT GCG	201
Lys Phe, Pro Arg Ser Arg Pro Asn Pro Ile Leu Arg Cys Ser Ile Ala	





TTA TCT TGG AAG CTT TCA AGT ATT AGT AAA CTG GAT AGT GGA GAG TAC Leu Ser Trp Lys Leu Ser Ser Ile Ser Lys Leu Asp Ser Gly Glu Tyr 310 315 320	1017
AGT TTG ACA TAT GAA ACA CCA GAA GGA GTG GTT TCT TTG CAG TGC AAA Ser Leu Thr Tyr Glu Thr Pro Glu Gly Val Val Ser Leu Gln Cys Lys 325 330 335	1065
ACT GTT GTC CTG ACC ATT CCT TCC TAT GTT GCT AGT ACA TTG CTG CGT Thr Val Val Leu Thr Ile Pro Ser Tyr Val Ala Ser Thr Leu Leu Arg 340 345 350	1113
CCT CTG TCT GCT GCT GCT GCA GAT GCA CTT TCA AAG TTT TAT TAC CCT Pro Leu Ser Ala Ala Ala Ala Asp Ala Leu Ser Lys Phe Tyr Tyr Pro 355 360 365	1161
CCA GTT GCT GCA GTT TCC ATA TCC TAT CCA AAA GAA GCT ATT AGA TCA Pro Val Ala Ala Val Ser Ile Ser Tyr Pro Lys Glu Ala Ile Arg Ser 370 375 380 385	1209
GAA TGC TTG ATA GAT GGT GAG TTG AAG GGG TTT GGT CAA TTG CAT CCA Glu Cys Leu Ile Asp Gly Glu Leu Lys Gly Phe Gly Gln Leu His Pro 390 395 400	1257
CGT AGC CAA GGA GTG GAA ACA TTA GGA ACT ATA TAC AGC TCA TCA CTA Arg Ser Gln Gly Val Glu Thr Leu Gly Thr Ile Tyr Ser Ser Ser Leu 405 410 415	1305
TTC CCC AAC CGA GCA CCA CCT GGA AGG GTT CTA CTC TTG AAT TAC ATT Phe Pro Asn Arg Ala Pro Pro Gly Arg Val Leu Leu Leu Asn Tyr Ile 420 425 430	1353
GGA GGA GCA ACT AAT ACT GGA ATT TTA TCG AAG ACG GAC AGT GAA CTT Gly Gly Ala Thr Asn Thr Gly Ile Leu Ser Lys Thr Asp Ser Glu Leu 435 440 445	1401
GTG GAA ACA GTT GAT CGA GAT TTG AGG AAA ATC CTT ATA AAC CCA AAT Val Glu Thr Val Asp Arg Asp Leu Arg Lys Ile Leu Ile Asn Pro Asn 450 455 460 465	1449
GCC CAG GAT CCA TTT GTA GTG GGG GTG AGA CTG TGG CCT CAA GCT ATT Ala Gln Asp Pro Phe Val Val Gly Val Arg Leu Trp Pro Gln Ala Ile 470 475 480	1497
CCA CAG TTC TTA GTT GGC CAT CTT GAT CTT CTA GAT GTT GCT AAA GCT Pro Gln Phe Leu Val Gly His Leu Asp Leu Leu Asp Val Ala Lys Ala 485 490 495	1545
TCT ATC AGA AAT ACT GGG TTT GAA GGG CTC TTC CTT GGG GGT AAT TAT Ser Ile Arg Asn Thr Gly Phe Glu Gly Leu Phe Leu Gly Gly Asn Tyr 500 505 510	1593
GTG TCT GGT GTT GCC TTG GGA CGA TGC GTT GAG GGA GCC TAT GAG GTA Val Ser Gly Val Ala Leu Gly Arg Cys Val Glu Gly Ala Tyr Glu Val 515 520 525	1641
GCA GCT GAA GTA AAC GAT TTT CTC ACA AAT AGA GTG TAC AAA Ala Ala Glu Val Asn Asp Phe Leu Thr Asn Arg Val Tyr Lys 530 535 540	1683
TAGTAGCAGT TTTTGT TTTT GTGGTGGAAT GGGTGATGGG ACTCTCGTGT TCCATTGAAT	1743
TATAATAATG TGAAAGTTTC TCAAATTCGT TCGATAGGTT TTTGGCGGCT TCTATTGCTG	1803

ATAATGTAAA ATCCTCTTTA AGTTTGAAAA AAAAAAAAAA AAAA

1847

## (2) INFORMATION FOR SEQ ID NO:12:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 543 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Met Val Ser Val Phe Asn Glu Ile Leu Phe Pro Pro Asn Gln Thr Leu  
 1 5 10 15  
 Leu Arg Pro Ser Leu His Ser Pro Thr Ser Phe Phe Thr Ser Pro Thr  
 20 25 30  
 Arg Lys Phe Pro Arg Ser Arg Pro Asn Pro Ile Leu Arg Cys Ser Ile  
 35 40 45  
 Ala Glu Glu Ser Thr Ala Ser Pro Pro Lys Thr Arg Asp Ser Ala Pro  
 50 55 60  
 Val Asp Cys Val Val Val Gly Gly Gly Val Ser Gly Leu Cys Ile Ala  
 65 70 75 80  
 Gln Ala Leu Ala Thr Lys His Ala Asn Ala Asn Val Val Val Thr Glu  
 85 90 95  
 Ala Arg Asp Arg Val Gly Gly Asn Ile Thr Thr Met Glu Arg Asp Gly  
 100 105 110  
 Tyr Leu Trp Glu Glu Gly Pro Asn Ser Phe Gln Pro Ser Asp Pro Met  
 115 120 125  
 Leu Thr Met Val Val Asp Ser Gly Leu Lys Asp Glu Leu Val Leu Gly  
 130 135 140  
 Asp Pro Asp Ala Pro Arg Phe Val Leu Trp Asn Arg Lys Leu Arg Pro  
 145 150 155 160  
 Val Pro Gly Lys Leu Thr Asp Leu Pro Phe Phe Asp Leu Met Ser Ile  
 165 170 175  
 Gly Gly Lys Ile Arg Ala Gly Phe Gly Ala Leu Gly Ile Arg Pro Pro  
 180 185 190  
 Pro Pro Gly His Glu Glu Ser Val Glu Glu Phe Val Arg Arg Asn Leu  
 195 200 205  
 Gly Asp Glu Val Phe Glu Arg Leu Ile Glu Pro Phe Cys Ser Gly Val  
 210 215 220  
 Tyr Ala Gly Asp Pro Ser Lys Leu Ser Met Lys Ala Ala Phe Gly Lys  
 225 230 235 240  
 Val Trp Lys Leu Glu Lys Asn Gly Gly Ser Ile Ile Gly Gly Thr Phe  
 245 250 255  
 Lys Ala Ile Gln Glu Arg Asn Gly Ala Ser Lys Pro Pro Arg Asp Pro  
 260 265 270

Arg Leu Pro Lys Pro Lys Gly Gln Thr Val Gly Ser Phe Arg Lys Gly  
 275 280 285  
 Leu Thr Met Leu Pro Asp Ala Ile Ser Ala Arg Leu Gly Asn Lys Val  
 290 295 300  
 Lys Leu Ser Trp Lys Leu Ser Ser Ile Ser Lys Leu Asp Ser Gly Glu  
 305 310 315 320  
 Tyr Ser Leu Thr Tyr Glu Thr Pro Glu Gly Val Val Ser Leu Gln Cys  
 325 330 335  
 Lys Thr Val Val Leu Thr Ile Pro Ser Tyr Val Ala Ser Thr Leu Leu  
 340 345 350  
 Arg Pro Leu Ser Ala Ala Ala Ala Asp Ala Leu Ser Lys Phe Tyr Tyr  
 355 360 365  
 Pro Pro Val Ala Ala Val Ser Ile Ser Tyr Pro Lys Glu Ala Ile Arg  
 370 375 380  
 Ser Glu Cys Leu Ile Asp Gly Glu Leu Lys Gly Phe Gly Gln Leu His  
 385 390 395 400  
 Pro Arg Ser Gln Gly Val Glu Thr Leu Gly Thr Ile Tyr Ser Ser Ser  
 405 410 415  
 Leu Phe Pro Asn Arg Ala Pro Pro Gly Arg Val Leu Leu Leu Asn Tyr  
 420 425 430  
 Ile Gly Gly Ala Thr Asn Thr Gly Ile Leu Ser Lys Thr Asp Ser Glu  
 435 440 445  
 Leu Val Glu Thr Val Asp Arg Asp Leu Arg Lys Ile Leu Ile Asn Pro  
 450 455 460  
 Asn Ala Gln Asp Pro Phe Val Val Gly Val Arg Leu Trp Pro Gln Ala  
 465 470 475 480  
 Ile Pro Gln Phe Leu Val Gly His Leu Asp Leu Leu Asp Val Ala Lys  
 485 490 495  
 Ala Ser Ile Arg Asn Thr Gly Phe Glu Gly Leu Phe Leu Gly Gly Asn  
 500 505 510  
 Tyr Val Ser Gly Val Ala Leu Gly Arg Cys Val Glu Gly Ala Tyr Glu  
 515 520 525  
 Val Ala Ala Glu Val Asn Asp Phe Leu Thr Asn Arg Val Tyr Lys  
 530 535 540

## (2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 583 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (ix) FEATURE:
  - (A) NAME/KEY: promoter

(B) LOCATION: 1..583  
 (D) OTHER INFORMATION: /function= "arabidopsis protox-1 promoter"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GAATTCCGAT CGAATTATAT AATTATCATA AATTTGAATA AGCATGTTGC CTTTTATTAA	60
AGAGGTTTAA TAAAGTTTGG TAATAATGGA CTTTGACTTC AAACCTCGATT CTCATGTAAT	120
TAATTAATAT TTACATCAAA ATTTGGTCAC TAATATTACC AAATTAATAT ACTAAAAATGT	180
TAATTGCGAA ATAAACACT AATTCCAAAT AAAGGGTCAT TATGATAAAC ACGTATTGAA	240
CTTGATAAAG CAAAGCAAAA ATAATGGGTT TCAAGGTTTG GGTATATAT GACAAAAAAA	300
AAAAAAGGTT TGGTTATATA TCTATTGGGC CTATAACCAT GTTATACAAA TTTGGGCCTA	360
ACTAAAATAA TAAATAAAC GTAATGGTCC TTTTATATT TGGGTCAAAC CCAACTCTAA	420
ACCCAAACCA AAGAAAAAGT ATACGGTACG GTACACAGAC TTATGGTGTG TGTGATTGCA	480
GGTGAATATT TCTCGTCGTC TTCTCCTTTC TTCTGAAGAA GATTACCCAA TCTGAAAAAA	540
ACCAAGAAGC TGACAAAATT CCGAATTCTC TGCGATTTC ATG	583

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3848 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(ix) FEATURE:

- (A) NAME/KEY: promoter
- (B) LOCATION: 1..3848
- (D) OTHER INFORMATION: /function= "maize protox-1 promoter"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

TCGATCTTTC TAGGCTGATC CCCAAATCTT CCTCCGAAGC CCCTGGCGCC TCTGCCCTT	60
GGAGCTGGTG GCCTGAAAGA GCTTTGCTGT TGCCCCGAAG ATTGTGAGGT ATATTGTGAC	120
CTCTGAGACT GACTTCCTTT GTCGTCACCT TGAGTGGAGT TATGGATTGA CCTGACGTGC	180
CTCAGATGGA TTCTTCCTCC GAAGCCCCTG GTCATTTCCG AGAATCTGTA ATCTTATTCC	240
CTTCTTTGGC GAAAATCTGT CAGCTTGGAT GACTCATCC ATCTTCTGAA GCAGCTTCTC	300
CAGAGTTTGT GGAGGCTTCC TGGCGAAATA TTGGGCTGTA GGTCTGGAC GAAGACCCTT	360
GATCATGGCC TCAATGACAA TCTCATTGGG CACCGTAGGC GCTTGTGCCC TCAATCGCAA	420
GAACCTTCGT ACATATGCCT GAAGGTATTC TTCGTGATCT TGTGTGCATT GGAACAGAGC	480
CTGAGCTGTG ACCGACTTCG TTTGAAAGCC TTGGAAGCTA GTAACCAACA TGTGCTTAAG	540

CTTCTGCCAC GACGTGATAG TCCCTGGCCG AAGAGAAGAA TACCATGTTT GGGCTACATT 600  
 CCGGACTGCC ATGACGAAGG ACTTCGCCAT GACTACAGTG TTGACCCCAT ACGAAGATAT 660  
 AGTTGCTTCG TAGCTCATCA GAAACTGCTT TGGATCTGAG TGCCCATCAT ACATGGGGAG 720  
 CTGAGGTGGC TTGTATGATG GGGGCCATGG GGTAGCCTGC AGTTCTGCTG CCAAGGGAGA 780  
 AGCATCATCA AAAGTAAAGG CATCATGATT AAAATCATCA TACCATCCAT CCTCGTTGAA 840  
 TAAGCCTTCT TGACGAAGCT CCCTGTGTTG GGGCCTTCGA TCTTGTTTCAT CTTGAACAAG 900  
 ATGACGCACT TCTTCAGTGG CTTTCGTCGAT CTTTCTTTGG AGATCAGCCA GTCGCACCAT 960  
 CTTCTCCTTC TTTCTTTGTA CTTGTTGATG GATGATCTCC ATGTCCCTGA TCTCTTGGTC 1020  
 CAACTCCTCC TCTTGGAGTG TCAGACTGGT GGCTTTCCCTC TTCCTGGCTTC GAGCCTCTCG 1080  
 AAGAGAAAGA GTTTCCTTGAT TTGGGTCCAG CGGCTGCAGT GCAGTGGTCC CTGGTGCTGA 1140  
 AGCTTTCCTC GGTGGCATGA CAAAGGTCAG TGCTTGCCGA AGGTGGTCGA AAAGGGTTCA 1200  
 CTAGAGGTGG GAGCCAATGT TGGGGACTTC TCAAGTGCTA TGAGTTAAGA ACAAGGCAAC 1260  
 ACAAAATGTT AAATATTAAT AGCTTTCATC TTTCGAAGCA TTATTTCCCT TTGGGTATAA 1320  
 TGATCTTCAG ACGAAAGAGT CCTTCATCAT TGCGATATAT GTTAATAGAA GGAGGAGCAT 1380  
 ATGAAATGTA AGAGACAACA TGAACAATCG GTAGCATTG TTAATTCATC ATCATTTTAT 1440  
 TATTATGGAA AAATAGAAAC AATATTGAAT TACAAATGTA CCTTTGGCTT GACAGAAGAT 1500  
 AAAAGTACAA GCTTGACGCA CGAGCAAGTA CAAGTCAGTG TGAACAGTAC GGGGGTACTG 1560  
 TTCATCTATT TATAGGCACA GGACACAGCC TGTGAGAAAT TACAGTCATG CCCTTTACAT 1620  
 TTACTATTGA CTTATAGAAA AATCTATGAG GACTGGATAG CCTTTTCCCC TTTAAGTCGG 1680  
 TGCCTTTTTC CGCGATTAAG CCGAATCTCC CTTGCGCATA GCTTCGGAGC ATCGGCAACC 1740  
 TTCGTCACGA TCATGCCCTT CTCATTGTGT ATGCTTTTAA TCCTGAATTC GAAGGTACCT 1800  
 GTCCATAAAC CATACTTGGA AGACATTGTT AAATTATGTT TTTGAGGACC TTCGGAGGAC 1860  
 GAAGGCCCCC AACAGTCGTG TTTTGTAGGA CCTTCGGAAG ATGAAGGCCC CCAACAAGAC 1920  
 CTATCCATAA AACCAACCTA TCCACAAAAC CGACCCCAT TACCCTTCAT TTGCCTCACC 1980  
 AACAAACCCTA ATTAGGTTGT TGGTTTAAAT TTTTtagggT CAATTGAGTC ATCACCATCC 2040  
 ACTGTCACTC CACAACTCA ATATCAATAA ACAGACTCAA TCACCCAAAC TGACCATACC 2100  
 CATAAAACCG CCCCACCCTT CTAGCGCCTC GCCAGAAACC AGAAACCCTG ATTCAGAGTT 2160  
 CAAACTTAAA ACGACCATAA CTTTCACCTT GGAACTCGAA TCAGGTCCAT TTTTTCCTAA 2220  
 ATCACACAAA ATTAAATTTT GCATCCGATA ATCAAGCCAT CTCTTCACTA TGGTTTAAAG 2280  
 TGTTGCTCAC ACTAGTGAT TATGGACTA ATCACCTGTG TATCTCATAC AATAACATAT 2340  
 CAGTACATCT AAGTTGTTAC TCAATTACCA AAACCGAATT ATAGCCTTCG AAAAAGGTAA 2400  
 TCGACTAGTC ACTCAATTAC CAAAATAAAA CTTTAGACTT TCATGTATGA CATCCAACAT 2460

GACACTGTAC TGGACTAAAC CACCTTTCAA GCTACACAAG GAGCAAAAAT AACTAATTTT 2520  
 CGTAGTTGTA GGAGCTAAAG TATATGTCCA CAACAATAGT TAAGGGAAGC CCCCAGGAC 2580  
 TTAAGAGTCC TTTTACCTCT TGAAACTTTT GTCGTGGTCT ACTTTTTCAC TTTAAACTTC 2640  
 AAAATTTGAC ATTTTATCAC CCCTTAACTC TTAAAACCAT TTAAATTACA TTCTTACTAG 2700  
 ATTATAGATG ATTTTGTTGT GAAAAGTTTT TAAGACATGT TTACACATTG ATTAAAATCA 2760  
 TTTGTTCAAT TTCCTAGAGT TAAATCTAAT CTTATTAAAA CTATTAGAGA TACTTTCACG 2820  
 AGCTCTAAAT ATTTTATTTT TTTCATTATG GAATTTTGTT AGAATTCTTA TAGACCTTTT 2880  
 TTTGTGGTTT AAAAGCCTTG CCATGTTTTT AACAAGTTTT TTTTCTATTT TTTGAAATTT 2940  
 TCTTGGAAC CACTTCTAAC CCGGTAGAAG ATTTATTTTG CTACACTTAT ATCTACAACA 3000  
 AAATCAACTT ATGAAATTGT CTTGGAACT ACCTCTAACC CGGTAGAATG AATTTGAATG 3060  
 AAAATTAAAC CAACTTACGG AATCGCCCAA CATATGTCGA TTAAAGTGGA TATGGATACA 3120  
 TATGAAGAAG CCCTAGAGAT AATCTAAATG GTTTCAGAAT TGAGGGTTAT TTTTGAAGT 3180  
 TTGATGGGAA GATAAGACCA TAACGGTAGT TCACAGAGAT AAAAGGGTTA TTTTTTTCAG 3240  
 AAATATTTGT GCTGCAATTG ATCCTGTGCC TCAAATTCAG CCTGCAACCA AGGCCAGGTT 3300  
 CTAGAGCGAA CAAGGCCAC GTCACCCGTG GCCCGTCAGG CGAAGCAGGT CTTGTGCAGA 3360  
 CTTTGAGAGG GATTGGATAT CAACGGAACC AATCACGCAC GGCAATGCGA TTCCAGCCC 3420  
 ACCTGTAACG TTCCAGTGGG CCATCCTTAA CTCCAAGCCC AACGGCCCTA CCCCATCTCG 3480  
 TCGTGTGATC CACTCCGCCG CACAGGCGCT CAGCTCCGCA ACGCCGCCGG AAATGGTGC 3540  
 CGCCACAGCC ACCGCCATGG CCACCGCTGC ATCGCCGCTA CTCAACGGGA CCCGAATACC 3600  
 TGC GCGGCTC CGCCATCGAG GACTCAGCGT GCGCTGCGCT GCTGTGGCGG GCGGCGGGC 3660  
 CGAGGCACCG GCATCCACCG GCGCGCGGCT GTCCGCGGAC TGC GTTGTGG TGGGCGGAGG 3720  
 CATCAGTGGC CTCTGCACCG CGCAGGCGCT GGCCACGCGG CACGGCGTCG GGGACGTGCT 3780  
 TGTCACGGAG GCCCGCGCCC GCCCGGCGG CAACATTACC ACCGTCGAGC GCCCGAGGA 3840  
 AGGGTACC 3848

## (2) INFORMATION FOR SEQ ID NO:15:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1826 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## (iii) HYPOTHETICAL: NO

## (iv) ANTI-SENSE: NO

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: *Gossypium hirsutum* (cotton)

## (vii) IMMEDIATE SOURCE:

(B) CLONE: pWDC-15 (NRRL B-21594)

## (ix) FEATURE:

(A) NAME/KEY: misc\_feature

(B) LOCATION: 31..1647

(D) OTHER INFORMATION: /product= "Cotton protox-1 coding sequence"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

CCTCTCGCTC	GCCTGGCCCC	ACCACCAATC	ATGACGGCTC	TAATCGACCT	TTCTCTTCTC	60
CGTTCCTCGC	CCTCCGTTTC	CCCTTTCTCC	ATACCCCACC	ACCAGCATCC	GGCCCGCTTT	120
CGTAAACCTT	TCAAGCTCCG	ATGCTCCCTC	GCCGAGGGTC	CCACGATTTC	CTCATCTAAA	180
ATCGACGGGG	GAGAATCATC	CATCGCGGAT	TGCGTCATCG	TTGGAGGTGG	TATCAGTGGA	240
CTTTGCATTG	CTCAAGCTCT	CGCCACCAAG	CACCGTGACG	TCGCTTCCAA	TGTGATTGTG	300
ACGGAGGCCA	GAGACCGTGT	TGGTGGCAAC	ATCACTACCG	TTGAGAGAGA	TGGATATCTG	360
TGGGAAGAAG	GGCCCAACAG	TTTTCAGCCC	TCCGATCCTA	TTCTAACCAT	GGCCGTGGAT	420
AGTGGATTGA	AGGACGATTT	GGTTTTAGGT	GACCCTAATG	CACCGCGATT	TGTACTATGG	480
GAGGGAAAAC	TAAGGCCGTG	GCCCTCCAAG	CCAACCGACT	TGCCGTTTTT	TGATTTGATG	540
AGCATTGCTG	GAAAACCTAG	GGCTGGGTTT	GGGGCTATTG	GCATTCGGCC	TCCCCCTCCG	600
GGTTATGAAG	AATCGGTGGA	GGAGTTTGTG	CGCCGTAATC	TTGGTGCTGA	GGTTTTTGAA	660
CGCTTTATTG	AACCATTTTG	TTCAGGTGTT	TATGCAGGGG	ATCCTTCAAA	ATTAAGCATG	720
AAAGCAGCAT	TTGGAAGAGT	ATGGAAGCTA	GAAGAGATTG	GTGGCAGCAT	CATTGGTGGC	780
ACTTTCAAGA	CAATCCAGGA	GAGAAATAAG	ACACCTAAGC	CACCCAGAGA	CCCgcGCTCTG	840
CCAAAACCGA	AGGGCCAAAC	AGTTGGATCT	TTTAGGAAGG	GACTTACCAT	GCTGCCTGAG	900
GCAATTGCTA	ACAGTTTGGG	TAGCAATGTA	AAATTATCTT	GGAAGCTTTC	CAGTATTACC	960
AAATTGGGCA	ATGGAGGGTA	TAAGTTGACA	TTTGAAACAC	CTGAAGGAAT	GGTATCTCTT	1020
CAGAGTAGAA	GTGTTGTAAT	GACCATTTCA	TCCCATGTTG	CCAGTAACTT	GTTGCATCCT	1080
CTCTCGGCTG	CTGCTGCAGA	TGCATTATCC	CAATTTTATT	ATCCTCCAGT	TGCATCAGTC	1140
ACAGTCTCCT	ATCCAAAAGA	AGCCATTCTG	AAAGAATGTT	TGATTGATGG	TGAACCTAAG	1200
GGGTTTGCC	AGTTGCACCC	ACGCAGCCAA	GGAATTGAAA	CTTTAGGGAC	GATATACAGT	1260
TCATCACTTT	TCCCCAATCG	AGCTCCATCT	GGCAGGGTGT	TGCTCTTGAA	CTACATAGGA	1320
GGAGCTACCA	ACACTGGAAT	TTTGTCCAAG	ACTGAAGGGG	AACTTGTAAG	AGCAGTTGAT	1380
CGTGATTGTA	GAAAAATGCT	TATAAATCCT	AATGCAAAGG	ATCCTCTTGT	TTTGGGTGTA	1440
AGAGTATGGC	CAAAAAGCCAT	TCCACAGTTC	TTGGTTGGTC	ATTTGGATCT	CCTTGATAGT	1500
GCAAAAATGG	CTCTCAGGGA	TTCTGGGTTT	CATGGACTGT	TTCTTGGGGG	CAACTATGTA	1560

TCTGGTGTGG CATTAGGACG GTGTGTGGAA GGTGCTTACG AGGTTGCAGC TGAAGTGAAG 1620  
 GAATTCCTGT CACAATATGC ATACAAATAA TATTGAAATT CTTGTCAGGC TGCAAATGTA 1680  
 GAAGTCAGTT ATTGGATAGT ATCTCTTTAG CTA AAAAATT GGGTAGGGTT TTTTTTGTTA 1740  
 GTTCCTTGAC CACTTTTTGG GGTTCATT AGAACTTCAT ATTTGTATAT CATGTTGCAA 1800  
 TATCAAAAAA AAAAAAAA AAAAAA 1826

## (2) INFORMATION FOR SEQ ID NO:16:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 539 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: not relevant

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Met Thr Ala Leu Ile Asp Leu Ser Leu Leu Arg Ser Ser Pro Ser Val  
 1 5 10 15  
 Ser Pro Phe Ser Ile Pro His His Gln His Pro Pro Arg Phe Arg Lys  
 20 25 30  
 Pro Phe Lys Leu Arg Cys Ser Leu Ala Glu Gly Pro Thr Ile Ser Ser  
 35 40 45  
 Ser Lys Ile Asp Gly Gly Glu Ser Ser Ile Ala Asp Cys Val Ile Val  
 50 55 60  
 Gly Gly Gly Ile Ser Gly Leu Cys Ile Ala Gln Ala Leu Ala Thr Lys  
 65 70 75 80  
 His Arg Asp Val Ala Ser Asn Val Ile Val Thr Glu Ala Arg Asp Arg  
 85 90 95  
 Val Gly Gly Asn Ile Thr Thr Val Glu Arg Asp Gly Tyr Leu Trp Glu  
 100 105 110  
 Glu Gly Pro Asn Ser Phe Gln Pro Ser Asp Pro Ile Leu Thr Met Ala  
 115 120 125  
 Val Asp Ser Gly Leu Lys Asp Asp Leu Val Leu Gly Asp Pro Asn Ala  
 130 135 140  
 Pro Arg Phe Val Leu Trp Glu Gly Lys Leu Arg Pro Val Pro Ser Lys  
 145 150 155 160  
 Pro Thr Asp Leu Pro Phe Phe Asp Leu Met Ser Ile Ala Gly Lys Leu  
 165 170 175  
 Arg Ala Gly Phe Gly Ala Ile Gly Ile Arg Pro Pro Pro Pro Gly Tyr  
 180 185 190  
 Glu Glu Ser Val Glu Glu Phe Val Arg Arg Asn Leu Gly Ala Glu Val  
 195 200 205  
 Phe Glu Arg Phe Ile Glu Pro Phe Cys Ser Gly Val Tyr Ala Gly Asp



210	215	220
Pro Ser Lys Leu Ser Met Lys Ala Ala Phe Gly Arg Val Trp Lys Leu		
225	230	235 240
Glu Glu Ile Gly Gly Ser Ile Ile Gly Gly Thr Phe Lys Thr Ile Gln		
	245	250 255
Glu Arg Asn Lys Thr Pro Lys Pro Pro Arg Asp Pro Arg Leu Pro Lys		
	260	265 270
Pro Lys Gly Gln Thr Val Gly Ser Phe Arg Lys Gly Leu Thr Met Leu		
	275	280 285
Pro Glu Ala Ile Ala Asn Ser Leu Gly Ser Asn Val Lys Leu Ser Trp		
	290	295 300
Lys Leu Ser Ser Ile Thr Lys Leu Gly Asn Gly Gly Tyr Asn Leu Thr		
	310	315 320
Phe Glu Thr Pro Glu Gly Met Val Ser Leu Gln Ser Arg Ser Val Val		
	325	330 335
Met Thr Ile Pro Ser His Val Ala Ser Asn Leu Leu His Pro Leu Ser		
	340	345 350
Ala Ala Ala Ala Asp Ala Leu Ser Gln Phe Tyr Tyr Pro Pro Val Ala		
	355	360 365
Ser Val Thr Val Ser Tyr Pro Lys Glu Ala Ile Arg Lys Glu Cys Leu		
	370	375 380
Ile Asp Gly Glu Leu Lys Gly Phe Gly Gln Leu His Pro Arg Ser Gln		
	385	390 395 400
Gly Ile Glu Thr Leu Gly Thr Ile Tyr Ser Ser Ser Leu Phe Pro Asn		
	405	410 415
Arg Ala Pro Ser Gly Arg Val Leu Leu Leu Asn Tyr Ile Gly Gly Ala		
	420	425 430
Thr Asn Thr Gly Ile Leu Ser Lys Thr Glu Gly Glu Leu Val Glu Ala		
	435	440 445
Val Asp Arg Asp Leu Arg Lys Met Leu Ile Asn Pro Asn Ala Lys Asp		
	450	455 460
Pro Leu Val Leu Gly Val Arg Val Trp Pro Lys Ala Ile Pro Gln Phe		
	465	470 475 480
Leu Val Gly His Leu Asp Leu Leu Asp Ser Ala Lys Met Ala Leu Arg		
	485	490 495
Asp Ser Gly Phe His Gly Leu Phe Leu Gly Gly Asn Tyr Val Ser Gly		
	500	505 510
Val Ala Leu Gly Arg Cys Val Glu Gly Ala Tyr Glu Val Ala Ala Glu		
	515	520 525
Val Lys Glu Phe Leu Ser Gln Tyr Ala Tyr Lys		
	530	535

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1910 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Beta vulgaris (Sugar Beet)

(vii) IMMEDIATE SOURCE:

(B) CLONE: pWDC-16 (NRRL B-21595N)

(ix) FEATURE:

(A) NAME/KEY: misc\_feature

(B) LOCATION: 1..1680

(D) OTHER INFORMATION: /product= "Sugar Beet protox-1 coding sequence"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

ATGAAATCAA TGGCGTTATC AAACATGCATT CCACAGACAC AGTGCATGCC ATTGCGCAGC	60
AGCGGGCATT ACAGGGGTAA TTGTATCATG TTGTCAATTC CATGTAGTTT AATTGGAAGA	120
CGAGGTTATT ATTCACATAA GAAGAGGAGG ATGAGCATGA GTTGCAGCAC AAGCTCAGGC	180
TCAAAGTCAG CGGTTAAAGA AGCAGGATCA GGATCAGGTG CAGGAGGATT GCTAGACTGC	240
GTAATCGTTG GAGGTGGAAT TAGCGGGCTT TGCATCGCGC AGGCTCTTTG TACAAAACAC	300
TCCTCTTCCT CTTTATCCCC AAATTTTATA GTTACAGAGG CCAAAGACAG AGTTGGCGGC	360
AACATCGTCA CTGTGGAGGC CGATGGCTAT ATCTGGGAGG AGGGACCCAA TAGCTTCCAG	420
CCTTCCGACG CGGTGCTCAC CATGGCGGTC GACAGTGGCT TGAAAGATGA GTTGGTGCTC	480
GGAGATCCCA ATGCTCCTCG CTTTGTGCTA TGGAATGACA AATTAAGGCC CGTACCTTCC	540
AGTCTCACCG ACCTCCCTTT CTTGACCTC ATGACCATTG CGGGCAAGAT TAGGGCTGCT	600
CTTGGTGCTC TCGGATTTG CCCTTCTCCT CCACCTCATG AGGAATCTGT TGAACACTTT	660
GTGCGTCGTA ATCTCGGAGA TGAGGTCTTT GAACGCTTGA TTGAACCTTT TTGTTTCAAGT	720
GTGTATGCCG GTGATCCTGC CAAGCTGAGT ATGAAAGCTG CTTTGTGGAA GGTCTGGAAG	780
TTGGAGCAAA AGGGTGGCAG CATAATTGGT GGCACCTCTCA AAGCTATACA GGAAAGAGGG	840
AGTAATCCTA AGCCGCCCCG TGACCAGCGC CTCCCTAAAC CAAAGGGTCA GACTGTTGGA	900
TCCTTTAGAA AGGGACTCGT TATGTTGCCT ACCGCCATTT CTGCTCGACT TGGCAGTAGA	960
GTGAAACTAT CTTGGACCCT TTCTAGTATC GTAAAGTCAC TCAATGGAGA ATATAGTCTG	1020
ACTTATGATA CCCAGATGG CTTGGTTTCT GTAAGAACCA AAAGTGTGTG GATGACTGTT	1080
CCATCATATG TTGCAAGTAG GCTTCTTCGT CCACTTTTCTG ACTCTGCTGC AGATTCTCTT	1140
TCAAAATTTT ACTATCCACC AGTTGCAGCA GTGTCACTTT CCTATCCTAA AGAAGCGATC	1200

AGATCAGAAT GCTTGATTAA TGGTGAACCTT CAAGGTTTCG GGCAACTACA TCCCCGCAGT 1260  
 CAGGGTGTGG AAACCTTGGG AACAAATTTAT AGTTCGTCTC TTTTCCCTGG TCGAGCACCA 1320  
 CCTGGTAGGA TCTTGATCTT GAGCTACATC GGAGGTGCTA AAAATCCTGG CATATTAAAC 1380  
 AAGTCGAAAG ATGAACTTGC CAAGACAGTT GACAAGGACC TGAGAAGAAT GCTTATAAAT 1440  
 CCTGATGCAA AACTTCCTCG TGTACTGGGT GTGAGAGTAT GGCCTCAAGC AATACCCCAG 1500  
 TTTTCTATTG GGCACCTTGA TCTGCTCGAT GCTGCAAAAG CTGCTCTGAC AGATACAGGG 1560  
 GTCAAAGGAC TGTTTCTTGG TGGCAACTAT GTTTCAGGTG TTGCCTTGGG GCGGTGTATA 1620  
 GAGGGTGCTT ATGAGTCTGC AGCTGAGGTA GTAGATTTCC TCTCACAGTA CTCAGACAAA 1680  
 TAGAGCTTCA GCATCCTGTG TAATTCAACA CAGGCCTTTT TGTATCTGTT GTGCGCGCAT 1740  
 GTAGTCTGGT CGTGGTGCTA GGATTGATTA GTTGCTCTGC TGTGTGATCC ACAAGAATTT 1800  
 TGATGGAATT TTTCCAGATG TGGGCATTAT ATGTTGCTGT CTTATAAATC CTTAATTTGT 1860  
 ACGTTTAGTG AATTACACCG CATTTGATGA CTAAAAAAA AAAAAAAA 1910

## (2) INFORMATION FOR SEQ ID NO:18:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 560 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: not relevant

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Met Lys Ser Met Ala Leu Ser Asn Cys Ile Pro Gln Thr Gln Cys Met  
 1 5 10 15  
 Pro Leu Arg Ser Ser Gly His Tyr Arg Gly Asn Cys Ile Met Leu Ser  
 20 25 30  
 Ile Pro Cys Ser Leu Ile Gly Arg Arg Gly Tyr Tyr Ser His Lys Lys  
 35 40 45  
 Arg Arg Met Ser Met Ser Cys Ser Thr Ser Ser Gly Ser Lys Ser Ala  
 50 55 60  
 Val Lys Glu Ala Gly Ser Gly Ser Gly Ala Gly Gly Leu Leu Asp Cys  
 65 70 75 80  
 Val Ile Val Gly Gly Gly Ile Ser Gly Leu Cys Ile Ala Gln Ala Leu  
 85 90 95  
 Cys Thr Lys His Ser Ser Ser Ser Leu Ser Pro Asn Phe Ile Val Thr  
 100 105 110  
 Glu Ala Lys Asp Arg Val Gly Gly Asn Ile Val Thr Val Glu Ala Asp  
 115 120 125  
 Gly Tyr Ile Trp Glu Glu Gly Pro Asn Ser Phe Gln Pro Ser Asp Ala

130	135	140
Val Leu Thr Met Ala 145	Val Asp Ser Gly Leu 150	Lys Asp Glu Leu Val Leu 155 160
Gly Asp Pro Asn Ala 165	Pro Arg Phe Val 170	Leu Trp Asn Asp Lys Leu Arg 175
Pro Val Pro Ser Ser 180	Leu Thr Asp Leu 185	Pro Phe Phe Asp Leu Met Thr 190
Ile Pro Gly Lys Ile 195	Arg Ala Ala Leu 200	Gly Ala Leu Gly Phe Arg Pro 205
Ser Pro Pro Pro His 210	Glu Glu Ser Val 215	Glu His Phe Val Arg Arg Asn 220
Leu Gly Asp Glu Val 225	Phe Glu Arg Leu 230	Ile Glu Pro Phe Cys Ser Gly 235 240
Val Tyr Ala Gly Asp 245	Pro Ala Lys Leu 250	Ser Met Lys Ala Ala Phe Gly 255
Lys Val Trp Lys Leu 260	Glu Gln Lys Gly 265	Gly Ser Ile Ile Gly Gly Thr 270
Leu Lys Ala Ile Gln 275	Glu Arg Gly Ser 280	Asn Pro Lys Pro Pro Arg Asp 285
Gln Arg Leu Pro Lys 290	Pro Lys Gly Gln 295	Thr Val Gly Ser Phe Arg Lys 300
Gly Leu Val Met Leu 305	Pro Thr Ala Ile 310	Ser Ala Arg Leu Gly Ser Arg 315 320
Val Lys Leu Ser Trp 325	Thr Leu Ser Ser 330	Ile Val Lys Ser Leu Asn Gly 335
Glu Tyr Ser Leu Thr 340	Tyr Asp Thr Pro 345	Asp Gly Leu Val Ser Val Arg 350
Thr Lys Ser Val Val 355	Met Thr Val Pro 360	Ser Tyr Val Ala Ser Arg Leu 365
Leu Arg Pro Leu Ser 370	Asp Ser Ala Ala 375	Asp Ser Leu Ser Lys Phe Tyr 380
Tyr Pro Pro Val Ala 385	Ala Val Ser Leu 390	Ser Tyr Pro Lys Glu Ala Ile 395 400
Arg Ser Glu Cys Leu 405	Ile Asn Gly Glu 410	Leu Gln Gly Phe Gly Gln Leu 415
His Pro Arg Ser Gln 420	Gly Val Glu Thr 425	Leu Gly Thr Ile Tyr Ser Ser 430
Ser Leu Phe Pro Gly 435	Arg Ala Pro Pro 440	Gly Arg Ile Leu Ile Leu Ser 445
Tyr Ile Gly Gly Ala 450	Lys Asn Pro Gly 455	Ile Leu Asn Lys Ser Lys Asp 460
Glu Leu Ala Lys Thr 465	Val Asp Lys Asp 470	Leu Arg Arg Met Leu Ile Asn 475 480

```

Pro Asp Ala Lys Leu Pro Arg Val Leu Gly Val Arg Val Trp Pro Gln
      485                      490                      495

Ala Ile Pro Gln Phe Ser Ile Gly His Phe Asp Leu Leu Asp Ala Ala
      500                      505                      510

Lys Ala Ala Leu Thr Asp Thr Gly Val Lys Gly Leu Phe Leu Gly Gly
      515                      520                      525

Asn Tyr Val Ser Gly Val Ala Leu Gly Arg Cys Ile Glu Gly Ala Tyr
      530                      535                      540

Glu Ser Ala Ala Glu Val Val Asp Phe Leu Ser Gln Tyr Ser Asp Lys
      545                      550                      555                      560

```

## (2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1784 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: Brassica napus (oilseed rape)
- (vii) IMMEDIATE SOURCE:
  - (B) CLONE: pWDC-17 (NRRL B-21615)
- (ix) FEATURE:
  - (A) NAME/KEY: misc\_feature
  - (B) LOCATION: 47..1654
  - (D) OTHER INFORMATION: /product= "Oilseed rape protox-1 coding sequence"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

```

GGGCCCCCCC CAAAATTGAG GATTCTCCTT CTCGCGGGCG ATCGCCATGG ATTTATCTCT      60
TCTCCGTCCG CAGCCATTCC TATCGCCATT CTCAAATCCA TTTCCTCGGT CGCGTCCCTA      120
CAAGCCTCTC AACCTCCGTT GCTCCGTATC CGGTGGATCC GTCGTCGGCT CTTCTACAAT      180
CGAAGGCGGA GGAGGAGGTA AAACCGTCAC GGCGGACTGC GTGATCGTCG GCGGAGGAAT      240
CAGCGGCCTG TGCATTGCGC AAGCGCTCGT GACGAAGCAC CCAGACGCTG CAAAGAATGT      300
GATGGTGACG GAGGCGAAGG ACCGTGTGGG AGGGAATATC ATCACGCGAG AGGAGCAAGG      360
GTTTCTATGG GAAGAAGGTC CCAATAGCTT TCAGCCGTCT GATCCTATGC TCACTATGGT      420
GGTAGATAGT GGTTTGAAAG ATGATCTAGT CTTGGGAGAT CCTACTGCTC CGAGGTTTGT      480
GTTGTGGAAT GGAAGCTGA GGCCGGTTCC GTCGAAGCTA ACTGACTTGC CTTTCTTTGA      540
CTTGATGAGT ATTGGAGGGA AGATTAGAGC TGGGTTTGGT GCCATTGGTA TTCGACCTTC      600

```

ACCTCCGGGT CGTGAGGAAT CAGTGGAAGA GTTTGTAAGG CGTAATCTTG GTGATGAGGT 660  
 TTTTGAGCGC TTGATTGAAC CCTTTTGCTC AGGTGTTTAT GCGGGAGATC CTGCGAAACT 720  
 GAGTATGAAA GCAGCTTTTG GGAAGGTTTG GAAGCTAGAG GAGAATGGTG GGAGCATCAT 780  
 TGGTGGTGCT TTTAAGGCAA TTCAAGCGAA AAATAAAGCT CCCAAGACAA CCCGAGATCC 840  
 GCGTCTGCCA AAGCCAAAGG GCCAAACTGT TGGTTCCTTC AGGAAAGGAC TCACAATGCT 900  
 GCCAGAGGCA ATCTCCGCAA GGTGGGTGA CAAGGTGAAA GTTCTTGGA AGCTCTCAAG 960  
 TATCACTAAG CTGGCCAGCG GAGAATATAG CTTAACTTAC GAAACTCCGG AGGGTATAGT 1020  
 CACTGTACAG AGCAAAAAGTG TAGTGATGAC TGTGCCATCT CATGTTGCTA GTAGTCTCTT 1080  
 GCGCCCTCTC TCTGATTCTG CAGCTGAAGC GCTCTCAAAA CTCTACTATC CGCCAGTTGC 1140  
 AGCCGTATCC ATCTCATACG CGAAAGAAGC AATCCGAAGC GAATGCTTAA TAGATGGTGA 1200  
 ACTAAAAGGG TTCGGCCAGT TGCATCCACG CACGCAAAAA GTGGAAACTC TTGGAACAAT 1260  
 ATACAGTTCA TCGCTCTTTC CCAACCGAGC ACCGCCTGGA AGAGTATTGC TATTGAACTA 1320  
 CATCGGTGGA GCTACCAACA CTGGGATCTT ATCAAAGTCG GAAGGTGAGT TAGTGGAAGC 1380  
 AGTAGATAGA GACTTGAGGA AGATGCTGAT AAAGCCAAGC TCGACCGATC CACTTGTAAT 1440  
 TGGAGTAAAA TTATGGCCTC AAGCCATTCC TCAGTTTCTG ATAGGTCACA TTGATTTGGT 1500  
 AGACGCAGCG AAAGCATCGC TCTCGTCATC TGGTCATGAG GGCTTATTCT TGGGTGGAAA 1560  
 TTACGTTGCC GGTGTAGCAT TGGGTCGGTG TGTGGAAGGT GCTTATGAAA CTGCAACCCA 1620  
 AGTGAATGAT TTCATGTCAA GGTATGCTTA CAAGTAATGT AACGCAGCAA CGATTTGATA 1680  
 CTAAGTAGTA GATTTTGCAG TTTTGACTTT AAGAACACTC TGTTTGTGAA AAATTCAAGT 1740  
 CTGTGATTGA GTAAATTTAT GTATTATTAC TAAAAA AAAA 1784

## (2) INFORMATION FOR SEQ ID NO:20:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 536 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Met Asp Leu Ser Leu Leu Arg Pro Gln Pro Phe Leu Ser Pro Phe Ser  
 1 5 10 15  
 Asn Pro Phe Pro Arg Ser Arg Pro Tyr Lys Pro Leu Asn Leu Arg Cys  
 20 25 30  
 Ser Val Ser Gly Gly Ser Val Val Gly Ser Ser Thr Ile Glu Gly Gly  
 35 40 45  
 Gly Gly Gly Lys Thr Val Thr Ala Asp Cys Val Ile Val Gly Gly Gly

50

55

60

Ile Ser Gly Leu Cys Ile Ala Gln Ala Leu Val Thr Lys His Pro Asp  
 65 70 75 80  
 Ala Ala Lys Asn Val Met Val Thr Glu Ala Lys Asp Arg Val Gly Gly  
 85 90 95  
 Asn Ile Ile Thr Arg Glu Glu Gln Gly Phe Leu Trp Glu Glu Gly Pro  
 100 105 110  
 Asn Ser Phe Gln Pro Ser Asp Pro Met Leu Thr Met Val Val Asp Ser  
 115 120 125  
 Gly Leu Lys Asp Asp Leu Val Leu Gly Asp Pro Thr Ala Pro Arg Phe  
 130 135 140  
 Val Leu Trp Asn Gly Lys Leu Arg Pro Val Pro Ser Lys Leu Thr Asp  
 145 150 155 160  
 Leu Pro Phe Phe Asp Leu Met Ser Ile Gly Gly Lys Ile Arg Ala Gly  
 165 170 175  
 Phe Gly Ala Ile Gly Ile Arg Pro Ser Pro Pro Gly Arg Glu Glu Ser  
 180 185 190  
 Val Glu Glu Phe Val Arg Arg Asn Leu Gly Asp Glu Val Phe Glu Arg  
 195 200 205  
 Leu Ile Glu Pro Phe Cys Ser Gly Val Tyr Ala Gly Asp Pro Ala Lys  
 210 215 220  
 Leu Ser Met Lys Ala Ala Phe Gly Lys Val Trp Lys Leu Glu Glu Asn  
 225 230 235 240  
 Gly Gly Ser Ile Ile Gly Gly Ala Phe Lys Ala Ile Gln Ala Lys Asn  
 245 250 255  
 Lys Ala Pro Lys Thr Thr Arg Asp Pro Arg Leu Pro Lys Pro Lys Gly  
 260 265 270  
 Gln Thr Val Gly Ser Phe Arg Lys Gly Leu Thr Met Leu Pro Glu Ala  
 275 280 285  
 Ile Ser Ala Arg Leu Gly Asp Lys Val Lys Val Ser Trp Lys Leu Ser  
 290 295 300  
 Ser Ile Thr Lys Leu Ala Ser Gly Glu Tyr Ser Leu Thr Tyr Glu Thr  
 305 310 315 320  
 Pro Glu Gly Ile Val Thr Val Gln Ser Lys Ser Val Val Met Thr Val  
 325 330 335  
 Pro Ser His Val Ala Ser Ser Leu Leu Arg Pro Leu Ser Asp Ser Ala  
 340 345 350  
 Ala Glu Ala Leu Ser Lys Leu Tyr Tyr Pro Pro Val Ala Ala Val Ser  
 355 360 365  
 Ile Ser Tyr Ala Lys Glu Ala Ile Arg Ser Glu Cys Leu Ile Asp Gly  
 370 375 380  
 Glu Leu Lys Gly Phe Gly Gln Leu His Pro Arg Thr Gln Lys Val Glu  
 385 390 395 400

005021-6250E260

Thr Leu Gly Thr Ile Tyr Ser Ser Ser Leu Phe Pro Asn Arg Ala Pro  
 405 410 415  
 Pro Gly Arg Val Leu Leu Leu Asn Tyr Ile Gly Gly Ala Thr Asn Thr  
 420 425 430  
 Gly Ile Leu Ser Lys Ser Glu Gly Glu Leu Val Glu Ala Val Asp Arg  
 435 440 445  
 Asp Leu Arg Lys Met Leu Ile Lys Pro Ser Ser Thr Asp Pro Leu Val  
 450 455 460  
 Leu Gly Val Lys Leu Trp Pro Gln Ala Ile Pro Gln Phe Leu Ile Gly  
 465 470 475 480  
 His Ile Asp Leu Val Asp Ala Ala Lys Ala Ser Leu Ser Ser Ser Gly  
 485 490 495  
 His Glu Gly Leu Phe Leu Gly Gly Asn Tyr Val Ala Gly Val Ala Leu  
 500 505 510  
 Gly Arg Cys Val Glu Gly Ala Tyr Glu Thr Ala Thr Gln Val Asn Asp  
 515 520 525  
 Phe Met Ser Arg Tyr Ala Tyr Lys  
 530 535

## (2) INFORMATION FOR SEQ ID NO:21:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1224 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: Oryza sativa (rice)
- (vii) IMMEDIATE SOURCE:
  - (B) CLONE: pWDC-18 (NRRL B-21648)
- (ix) FEATURE:
  - (A) NAME/KEY: misc\_feature
  - (B) LOCATION: 1..936
  - (D) OTHER INFORMATION: /product= "Rice protox-1 partial coding sequence"
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:
 

CGGGCTTTGA AGGCTGCATT TGGGAAGGTG TGGAGGCTGG AGGATACTGG AGGTAGCATT	60
ATTGGTGGAA CCATCAAGAC AATCCAGGAG AGGGGGAAAA ACCCCAAACC GCCGAGGGAT	120
CCCCGCCTTC CAACGCCAAA GGGGCAGACA GTTGCATCTT TCAGGAAGGG TCTGACTATG	180
CTCCCGGATG CTATTACATC TAGGTTGGGT AGCAAAGTCA AACTTTCATG GAAGTTGACA	240
AGCATTACAA AGTCAGACAA CAAAGGATAT GCATTAGTGT ATGAAACACC AGAAGGGGTG	300



GTCTCGGTGC AAGCTAAAAC TGTGTGCATG ACCATCCCAT CATATGTTGC TAGTGATATC 360  
 TTGCGGCCAC TTTCAAGTGA TGCAGCAGAT GCTCTGTCAA TATTCTATTA TCCACCAGTT 420  
 GCTGCTGTAA CTGTTTCATA TCCAAAAGAA GCAATTAGAA AAGAATGCTT AATTGACGGA 480  
 GAGCTCCAGG GTTTCGGCCA GCTGCATCCG CGTAGTCAGG GAGTTGAGAC TTTAGGAACA 540  
 ATATATAGCT CATCACTCTT TCCAAATCGT GCTCCAGCTG GAAGGGTGTT ACTTCTGAAC 600  
 TACATAGGAG GTTCTACAAA TACAGGGATT GTTTCCAAGA CTGAAAGTGA GCTGGTAGAA 660  
 GCAGTTGACC GTGACCTCAG GAAGATGCTG ATAAATCCTA GAGCAGTGGA CCCTTTGGTC 720  
 CTTGGCGTCC GGGTATGGCC ACAAGCCATA CCACAGTTCC TCATTGGCCA TCTTGATCAT 780  
 CTTGAGGCTG CAAAATCTGC CCTGGGCAAA GGTGGGTATG ATGGATTGTT CCTCGGAGGG 840  
 AACTATGTTG CAGGAGTTGC CCTGGGCCGA TCGTGTGAAG GTGCATATGA GAGTGCCTCA 900  
 CAAATATCTG ACTACTTGAC CAAGTACGCC TACAAGTGAT CAAAGTTGGC CTGCTCCTTT 960  
 TGGCACATAG ATGTGAGGCT TCTAGCAGCA AAAATTTTCAT GGGCATCTTT TTATCCTGAT 1020  
 TCTAATTAGT TAGAATTTAG AATTGTAGAG GAATGTTCCA TTTGCAGTTC ATAATAGTTG 1080  
 TTCAGATTTT AGCCATTCAA TTTGTGCAGC CATTTACTAT ATGTAGTATG ATCTTGTAAG 1140  
 TACTACTAAG AACAAATCAA TTATATTTTC CTGCAAGTGA CATCTTAATC GTCAGCAAAT 1200  
 CCAGTTACTA GTAAAAA AAAA 1224

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 312 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

Arg	Ala	Leu	Lys	Ala	Ala	Phe	Gly	Lys	Val	Trp	Arg	Leu	Glu	Asp	Thr
1			5						10					15	
Gly	Gly	Ser	Ile	Ile	Gly	Gly	Thr	Ile	Lys	Thr	Ile	Gln	Glu	Arg	Gly
			20					25					30		
Lys	Asn	Pro	Lys	Pro	Pro	Arg	Asp	Pro	Arg	Leu	Pro	Thr	Pro	Lys	Gly
			35				40					45			
Gln	Thr	Val	Ala	Ser	Phe	Arg	Lys	Gly	Leu	Thr	Met	Leu	Pro	Asp	Ala
		50				55					60				
Ile	Thr	Ser	Arg	Leu	Gly	Ser	Lys	Val	Lys	Leu	Ser	Trp	Lys	Leu	Thr
65				70					75					80	
Ser	Ile	Thr	Lys	Ser	Asp	Asn	Lys	Gly	Tyr	Ala	Leu	Val	Tyr	Glu	Thr
				85					90					95	

Pro Glu Gly Val Val Ser Val Gln Ala Lys Thr Val Val Met Thr Ile  
 100 105 110  
 Pro Ser Tyr Val Ala Ser Asp Ile Leu Arg Pro Leu Ser Ser Asp Ala  
 115 120 125  
 Ala Asp Ala Leu Ser Ile Phe Tyr Tyr Pro Pro Val Ala Ala Val Thr  
 130 135 140  
 Val Ser Tyr Pro Lys Glu Ala Ile Arg Lys Glu Cys Leu Ile Asp Gly  
 145 150 155 160  
 Glu Leu Gln Gly Phe Gly Gln Leu His Pro Arg Ser Gln Gly Val Glu  
 165 170 175  
 Thr Leu Gly Thr Ile Tyr Ser Ser Ser Leu Phe Pro Asn Arg Ala Pro  
 180 185 190  
 Ala Gly Arg Val Leu Leu Leu Asn Tyr Ile Gly Gly Ser Thr Asn Thr  
 195 200 205  
 Gly Ile Val Ser Lys Thr Glu Ser Glu Leu Val Glu Ala Val Asp Arg  
 210 215 220  
 Asp Leu Arg Lys Met Leu Ile Asn Pro Arg Ala Val Asp Pro Leu Val  
 225 230 235 240  
 Leu Gly Val Arg Val Trp Pro Gln Ala Ile Pro Gln Phe Leu Ile Gly  
 245 250 255  
 His Leu Asp His Leu Glu Ala Ala Lys Ser Ala Leu Gly Lys Gly Gly  
 260 265 270  
 Tyr Asp Gly Leu Phe Leu Gly Gly Asn Tyr Val Ala Gly Val Ala Leu  
 275 280 285  
 Gly Arg Cys Val Glu Gly Ala Tyr Glu Ser Ala Ser Gln Ile Ser Asp  
 290 295 300  
 Tyr Leu Thr Lys Tyr Ala Tyr Lys  
 305 310

## (2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1590 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: Sorghum bicolor (sorghum)
- (vii) IMMEDIATE SOURCE:
  - (B) CLONE: pWDC-19 (NRRL B-21649)
- (ix) FEATURE:
  - (A) NAME/KEY: misc\_feature

(B) LOCATION: 1..1320  
 (D) OTHER INFORMATION: /product= "Sorghum protox-1 partial coding sequence"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

TCCACCGTCG AGCGCCCCGA GGAAGGGTAC CTCTGGGAGG AGGGTCCCAA CAGCTTCCAG	60
CCATCCGACC CCGTTCTCTC CATGGCCGTG GACAGCGGGC TGAAGGATGA CCTGGTTTTT	120
GGGGACCCCA ACGCGCCACG GTTCGTGCTG TGGGAGGGGA AGCTGAGGCC CGTGCCATCC	180
AAGCCCCCGC ACCTCCCGTT CTTCGATCTC ATGAGCATCC CTGGCAAGCT CAGGGCCGGT	240
CTCGGCGCGC TTGGCATCCG CCCGCCTGCT CCAGGCCGCG AGGAGTCAGT GGAGGAGTTT	300
GTGCGCCGCA ACCTCGGTGC TGAGGTCTTT GAGCGCCTAA TTGAGCCTTT CTGCTCAGGT	360
GTCTATGCTG GCGATCCTTC CAAGCTCAGT ATGAAGGCTG CATTTGGGAA GGTGTGGCGG	420
TTAGAAGAAG CTGGAGGTAG TATTATTGGT GGAACCATCA AGACGATTCA GGAGAGGGGC	480
AAGAATCCAA AACCACCGAG GGATCCCCGC CTTCCGAAGC CAAAAGGGCA GACAGTTGCA	540
TCTTTCAGGA AGGGTCTTGC CATGCTTCCA AATGCCATCA CATCCAGCTT GGGTAGTAAA	600
GTCAAACATAT CATGGAAACT CACGAGCATG ACAAATCAG ATGGCAAGGG GTATGTTTTG	660
GAGTATGAAA CACCAGAAGG GGTGTTTTTG GTGCAGGCTA AAAGTGTTAT CATGACCATT	720
CCATCATATG TTGCTAGCGA CATTTTGCGT CCACCTTCAG GTGATGCTGC AGATGTTCTA	780
TCAAGATTCT ATTATCCACC AGTTGCTGCT GTAAQGGTTT CGTATCCAAA GGAAGCAATT	840
AGAAAAGAAT GCTTAATTGA TGGGGAAGTC CAGGGTTTTG GCCAGTTGCA TCCACGTAGT	900
CAAGGAGTTG AGACATTAGG AACAAATATAC AGCTCATCAC TCTTTCCAAA TCGTGCTCCT	960
GCTGGTAGGG TGTTACTTCT AAACACATA GGAGGTGCTA CAAACACAGG AATTGTTTCC	1020
AAGACTGAAA GTGAGCTGGT AGAAGCAGTT GACCGTGACC TCCGAAAAAT GCTTATAAAT	1080
CCTACAGCAG TGGACCCCTT AGTCCTTGGT GTCCGAGTTT GGCCACAAGC CATACTCAG	1140
TTCTTGGTAG GACATCTTGA TCTTCTGGAG GCCGCAAAT CTGCCCTGGA CCAAGGTGGC	1200
TATAATGGGC TGTTCCTAGG AGGGAACAT GTTGACAGGAG TTGCCCTGGG CAGATGCATT	1260
GAGGGCGCAT ATGAGAGTGC CGCGCAAATA TATGACTTCT TGACCAAGTA CGCCTACAAG	1320
TGATGGAAGA AGTGGAGCGC TGCTTGTTAA TTGTTATGTT GCATAGATGA GGTGAGACCA	1380
GGAGTAGTAA AAGGCGTCAC GAGTATTTTT CATTCTTATT TTGTAAATTG CACTTCTGTT	1440
TTTTTTTCCT GTCAGTAATT AGTTAGATTT TAGTTATGTA GGAGATTGTT GTGTTCACTG	1500
CCCTACAAAA GAATTTTAT TTTGCATTCG TTTATGAGAG CTGTGCAGAC TTATGTAACG	1560
TTTTACTGTA AGTATCAACA AAATCAAATA	1590

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 440 amino acids

- (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

```

Ser Thr Val Glu Arg Pro Glu Glu Gly Tyr Leu Trp Glu Glu Gly Pro
1          5          10          15
Asn Ser Phe Gln Pro Ser Asp Pro Val Leu Ser Met Ala Val Asp Ser
20          25          30
Gly Leu Lys Asp Asp Leu Val Phe Gly Asp Pro Asn Ala Pro Arg Phe
35          40          45
Val Leu Trp Glu Gly Lys Leu Arg Pro Val Pro Ser Lys Pro Ala Asp
50          55          60
Leu Pro Phe Phe Asp Leu Met Ser Ile Pro Gly Lys Leu Arg Ala Gly
65          70          75          80
Leu Gly Ala Leu Gly Ile Arg Pro Pro Ala Pro Gly Arg Glu Glu Ser
85          90          95
Val Glu Glu Phe Val Arg Arg Asn Leu Gly Ala Glu Val Phe Glu Arg
100         105         110
Leu Ile Glu Pro Phe Cys Ser Gly Val Tyr Ala Gly Asp Pro Ser Lys
115         120         125
Leu Ser Met Lys Ala Ala Phe Gly Lys Val Trp Arg Leu Glu Glu Ala
130         135         140
Gly Gly Ser Ile Ile Gly Gly Thr Ile Lys Thr Ile Gln Glu Arg Gly
145         150         155         160
Lys Asn Pro Lys Pro Pro Arg Asp Pro Arg Leu Pro Lys Pro Lys Gly
165         170         175
Gln Thr Val Ala Ser Phe Arg Lys Gly Leu Ala Met Leu Pro Asn Ala
180         185         190
Ile Thr Ser Ser Leu Gly Ser Lys Val Lys Leu Ser Trp Lys Leu Thr
195         200         205
Ser Met Thr Lys Ser Asp Gly Lys Gly Tyr Val Leu Glu Tyr Glu Thr
210         215         220
Pro Glu Gly Val Val Leu Val Gln Ala Lys Ser Val Ile Met Thr Ile
225         230         235         240
Pro Ser Tyr Val Ala Ser Asp Ile Leu Arg Pro Leu Ser Gly Asp Ala
245         250         255
Ala Asp Val Leu Ser Arg Phe Tyr Tyr Pro Pro Val Ala Ala Val Thr
260         265         270
Val Ser Tyr Pro Lys Glu Ala Ile Arg Lys Glu Cys Leu Ile Asp Gly
275         280         285

```

Glu Leu Gln Gly Phe Gly Gln Leu His Pro Arg Ser Gln Gly Val Glu  
 290 295 300  
 Thr Leu Gly Thr Ile Tyr Ser Ser Ser Leu Phe Pro Asn Arg Ala Pro  
 305 310 315 320  
 Ala Gly Arg Val Leu Leu Leu Asn Tyr Ile Gly Gly Ala Thr Asn Thr  
 325 330 335  
 Gly Ile Val Ser Lys Thr Glu Ser Glu Leu Val Glu Ala Val Asp Arg  
 340 345 350  
 Asp Leu Arg Lys Met Leu Ile Asn Pro Thr Ala Val Asp Pro Leu Val  
 355 360 365  
 Leu Gly Val Arg Val Trp Pro Gln Ala Ile Pro Gln Phe Leu Val Gly  
 370 375 380  
 His Leu Asp Leu Leu Glu Ala Ala Lys Ser Ala Leu Asp Gln Gly Gly  
 385 390 395 400  
 Tyr Asn Gly Leu Phe Leu Gly Gly Asn Tyr Val Ala Gly Val Ala Leu  
 405 410 415  
 Gly Arg Cys Ile Glu Gly Ala Tyr Glu Ser Ala Ala Gln Ile Tyr Asp  
 420 425 430  
 Phe Leu Thr Lys Tyr Ala Tyr Lys  
 435 440

## (2) INFORMATION FOR SEQ ID NO:25:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 93 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "maize protox-1 intron sequence"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

GTACGCTCCT CGCTGGCGCC GCAGCGTCTT CTTCTCAGAC TCATGCGCAG CCATGGAATT 60  
 GAGATGCTGA ATGGATTTTA TACGCGCGCG CAG 93

## (2) INFORMATION FOR SEQ ID NO:26:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2606 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (iii) HYPOTHETICAL: NO

## (iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:  
(A) ORGANISM: Beta vulgaris (sugar beet)

(vii) IMMEDIATE SOURCE:  
(B) CLONE: pWDC-20 (NRRL B-21650)

(ix) FEATURE:  
(A) NAME/KEY: misc\_feature  
(B) LOCATION: 2601..2606  
(D) OTHER INFORMATION: /note= "SalI site"

(ix) FEATURE:  
(A) NAME/KEY: misc\_feature  
(B) LOCATION: complement (1..538)  
(D) OTHER INFORMATION: /note= "partial cDNA of sugar beet  
protox-1"

(ix) FEATURE:  
(A) NAME/KEY: misc\_feature  
(B) LOCATION: 539..2606  
(D) OTHER INFORMATION: /note= "sugar beet protox-1  
promoter region (partial sequence of the ~ 3 kb PstI-SalI fragment subcloned  
from pWDC-20)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

CTGCAGGGGG AGGGAAAGAG AGACCGCGAC GGTGAGGGAG GGGAGACCGC GACGGTGAGG	60
GAGGGGAGAA CGCGACGGTG AGGGAGGGGA GAACGCGATG GTGAGGGAGG GGAGAACCGC	120
ACGCGCAGGG GAGGGGGATA ACTCGACGGT GCAGGGAGGT GAGGGGGACG ACGTGACGGC	180
GCAGGGGAGG GGGGAACCGT CGCGGGAAGG GGAAGACCGG GGGGCCGACA AGGTGGTGTT	240
ACTGGGGTAG GGAGAGGCGG CGTGGAGAAT AGTAACAGAG GGAGGAGTGG TGGTGCTAGG	300
GTGGAAGAAG GGTAAGAAAG AGGAAGAAAG AGAATTAACA TTATCTTAAC CAAACACCAC	360
TCTAAATCTA AGGGTTTTCT TTTCTTTTCC TCTCTCTCC CTTTCTTGAT TCCATTCCCT	420
TTACCCCGTT GCAACCAAAC GCCCCCTTAT TATGGACCGG AGGAAGTATG TAGAGATGGT	480
CACAAACTA CTTAAGCTGG TAACTTATAA ATATACTGGG TATTAAATGA ATTAAGTGCC	540
CACAAAATGA CTATAAATTA CTTTCGTAATC TTTAGGAACT ATGTTGGTCA CGAAATAACA	600
TAAACTGGT TATTTAATGG CTTTATGTAG GTACTGCATT CATAAATATA TTTCTAACAT	660
AATCGTGGTA TGTAGGTGTT TTATAACACA AGGATTAGGT TTACACCAAT GTCATTTTCA	720
TTAGAATGTA GTTAGAATCA CTTTGGAAC TTTGAAGAGTG ATGACACATT TTTATTATGC	780
TTTTATGAAA TGTCTTTGTG GTTTTTATGA TAGTATTGAG TTAAAGGCAA GTTGAAGTA	840
TATGATGGAG AAGTACAGTA TATAGGTGAC AATTGGTTTG CTTGTTTCTA TGAGTTGAAA	900
GATAAGTAGT ACACGACACT GAGCAATGAC CTCTTCTTAG TTGTAATTTT GTCTTCTCGA	960
CGTAGTGAAA GTACAAACAA GATTATGGCT TTCAAGCTTC CAAGATAACG AGATTGTATG	1020
AATTTTGTGG TGTATTTTAC ATCATTGTTT TACGTTGGAG ACAAATAAAA ACCAATGATG	1080
AGTTTGTGGA TTCGAGATTT GCCCCTAAGT CTTATTTACC CATGGCAAGC ATGCTGAAAC	1140

ATGTTAGTCA AAC TTACACA GCTACAATGT TTAGGGATTT TGAGCAAAAA ATTTGGGTAT 1200  
 TCTTTGGGTA CCATTATGTG AGTTGTTGAC TATGGATTAA ACAAATCAC TATATAAAGT 1260  
 CTGGAATGAG AAGCATCCGC AATTGACACA CCATGTTACT TTGATTGTTT CAACAAGTTT 1320  
 ATTAGATGTA TTTGTAGGAA TTTTGAAGAG GCGGAGATGT TGTGTTATAA TTGCTTTGGG 1380  
 GGTGCTTCAC ATGCACTCTG TTAGTGAGAC ATCTTCAGCT TATATTTTAA GGCGGTTAGT 1440  
 GAGTATGATT TTTTTTTTTC AAAC TTTTCG ATTTCCATGT AATTAAAAAA GGTGTTTGAT 1500  
 AAATACATGT TAAGATAGCC AAGAAAAGGC AAC TTTCAAA CAAATAAAAA AAATTAAGTC 1560  
 GCTTAATCAT TTTTCCAAGT ACTTTTACT TTTAACACCA CTTATTACTG AATCTATAGC 1620  
 CGTTAAGAAT GCATTTTCAC GCTCATACAT GCAAATCAAG AACCTCCTCA TTGAAGGAGA 1680  
 TAATTTAGTC CTCATAAACC CCGTTAAAGA CATTTT TAGC ATCCAGAGAA ATTTGATTC 1740  
 AGTTAAAATT GCATATATAA CCAGAGAAAC AAATTCAGAT GTTAGTCAGT CCAGCTACAT 1800  
 AGGTCAATGC CTGAGAGTTT AAAAGAATCC GTATCCTTAA GCATAAGTAG GTATTGAGGT 1860  
 GAGTTACAAA GGTAAGTTAC CGGTTACGCA CCACCTCCAC CAAACAAGTA TGGTTAGAAG 1920  
 ATACATGTAA TCGTTTATTT AGAGTACTAT TTATAAAAAA CTTTTTAACT AGAAACAGTT 1980  
 GTTTCATTTT GATATAAGGT TAATTAGAAT TCCCGAGCAA GCAAGAAGGG GATATAGAGG 2040  
 ATAAGGAGGG CGAGAGAGCG AGAGAGAGAT GAAATCAATG GCGTTATCAA ACTGCATTCC 2100  
 ACAGACACAG TGCATGCCAT TGCACAGCAG CGGGCATTAC AGGGGCAATT GTATCATGTT 2160  
 GTCAATTCCA TG TAGTTTAA TTGGAAGACG AGGTATTAT TCACATAAGA AGAGGAGGAT 2220  
 GAGCATGAGT TGCAGCACAA GCTCAGGCTC AAAGTCAGCG GTTAAAGAAG CAGGATCAGG 2280  
 ATCAGGATCA GGAGCAGGAG GATTGCTAGA CTGCGTAATC GTTGGAGGTG GAATTAGCGG 2340  
 GCTTTGCATC GCGCAGGCTC TTTGTACAAA ACAGTCCTCT TTATCCCCAA ATTTTATAGT 2400  
 GACAGAGGCC AAAGACAGAG TTGGCGGCAA CATCGTCACT GTGGAGGCCG ATGGCTATAT 2460  
 CTGGGAGGAG GGACCCAATA GCTTCAGCC TTCCGACGCG GTGCTCACCA TGGCGGTAAT 2520  
 TCTGTCTCTT CATTATTCAT AATCATAATT CAATTCAATT CAATTCCTAA CGTGGAATGT 2580  
 GGAATGTGGC ATGTGCGTAG GTCGAC 2606

## (2) INFORMATION FOR SEQ ID NO:27:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "Pclp\_Pla - plastid clpP gene promoter top strand PCR primer"

## (iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:

- (A) NAME/KEY: misc\_feature
- (B) LOCATION: 4..9
- (D) OTHER INFORMATION: /note= "EcoRI restriction site"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GCGGAATTCA TACTTATTTA TCATTAGAAA G

31

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "Pclp\_P1b - plastid clpP gene promoter bottom strand PCR primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:

- (A) NAME/KEY: misc\_feature
- (B) LOCATION: 4..9
- (D) OTHER INFORMATION: /note= "XbaI restriction site"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

GCGTCTAGAA AGAACTAAAT ACTATATTTTC AC

32

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "Pclp\_P2b - plastid clpP gene promoter bottom strand PCR primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:

- (A) NAME/KEY: misc\_feature
- (B) LOCATION: 4..9
- (D) OTHER INFORMATION: /note= "NcoI restriction site"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

005021"5250E/50



GCGCCATGGT AAATGAAAGA AAGAACTAAA

30

## (2) INFORMATION FOR SEQ ID NO:30:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 30 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: other nucleic acid
  - (A) DESCRIPTION: /desc = "Trps16\_P1a - plastid rps16 gene 3' untranslated region *XbaI/HindIII* top strand PCR primer"
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (ix) FEATURE:
  - (A) NAME/KEY: misc\_feature
  - (B) LOCATION: 4..9
  - (D) OTHER INFORMATION: /note= "*XbaI* restriction site"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GCGTCTAGAT CAACCGAAAT TCAATTAAGG

30

## (2) INFORMATION FOR SEQ ID NO:31:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 27 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: other nucleic acid
  - (A) DESCRIPTION: /desc = "Trps16\_p1b - plastid rps16 gene 3' untranslated region *XbaI/HindIII* bottom strand PCR primer"
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (ix) FEATURE:
  - (A) NAME/KEY: misc\_feature
  - (B) LOCATION: 4..9
  - (D) OTHER INFORMATION: /note= "*HindIII* restriction site"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

CGCAAGCTTC AATGGAAGCA ATGATAA

27

## (2) INFORMATION FOR SEQ ID NO:32:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 36 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid  
 (A) DESCRIPTION: /desc = "minpsb\_U - plastid psbA gene 5' untranslated region 38 nt (blunt/*NcoI*) including ATG start codon, top strand primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

GGGAGTCCCT GATGATTAAA TAAACCAAGA TTTTAC

36

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 40 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "minpsb\_L - plastid psbA gene 5' untranslated region 38 nt (blunt/*NcoI*) including ATG start codon (bottom strand primer)"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

CATGGTAAAA TCTTGGTTTA TTTAATCATC AGGGACTCCC

40

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "APRTXP1a - top strand PCR primer for amplifying the 5' portion of the mutant Arabidopsis protox gene"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:

- (A) NAME/KEY: misc\_feature
- (B) LOCATION: 5..10
- (D) OTHER INFORMATION: /note= "*NcoI* restriction site/ATG start codon"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

GGGACCATGG ATTGTGTGAT TGTCGGCGGA GG

32

## (2) INFORMATION FOR SEQ ID NO:35:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "APRTXP1b - bottom strand PCR primer for amplifying the 5' portion of the mutant Arabidopsis protox gene"

## (iii) HYPOTHETICAL: NO

## (iv) ANTI-SENSE: NO

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

CTCCGCTCTC CAGCTTAGTG ATAC

24

## (2) INFORMATION FOR SEQ ID NO:36:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 633 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: sugar cane

## (ix) FEATURE:

- (A) NAME/KEY: misc\_feature
- (B) LOCATION: 1..308
- (D) OTHER INFORMATION: /product= "Sugar cane protox-1 partial coding sequence"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

TTTCCAAGAC TGAAAGTGAG CTGGTAGAAG CAGTTGACCG TGACCTCCGG AAAATGCTTA	60
TAAATCCTAC AGCAGTGGAC CCTTTAGTCC TTGGTGTCCG AGTTTGGCCA CAAGCCATAC	120
CTCAGTTCCT GGTAGGACAT CTTGATCTTC TGGAGGCCGC AAAATCTGCC CTGGACCGAG	180
GTGGCTACGA TGGGCTGTTC CTAGGAGGGA ACTATGTTGC AGGAGTTGCC CTAGGCAGAT	240
GCGTTGAGGG CGCGTATGAG AGTGCCTCGC AAATATATGA CTTCTTGACC AAGTATGCCT	300
ACAAGTGATG AAAGAAGTGG AGTGCTGCTT GTTAATTGTT ATGTTGCATA GATGAGGTGA	360
GACCAGGAGT AGTAAAAGCG TTACGAGTAT TTTTCATTCT TATTTTGTA ATTGCACTTC	420
TGGTTTTTTC CTGTCAGTAA TTAGTTAGAT TTTAGTTCTG TAGGAGATTG TTCTGTTTAC	480

TGCCCTACAA AAGAATTTTT ATTTTGCATT CGTTTATGAG AGCTGTGCAG ACTTATGTAG 540  
 CGTTTTTCTG TAAGTACCAA CAAAATCAAA TACTATTCTG TAAGAGCTAA CAGAATGTGC 600  
 AACTGAGATT GCCTTGGATG AAAAAAAAAA AAA 633

## (2) INFORMATION FOR SEQ ID NO:37:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 101 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

Ser Lys Thr Glu Ser Glu Leu Val Glu Ala Val Asp Arg Asp Leu Arg  
 1 5 10 15  
 Lys Met Leu Ile Asn Pro Thr Ala Val Asp Pro Leu Val Leu Gly Val  
 20 25 30  
 Arg Val Trp Pro Gln Ala Ile Pro Gln Phe Leu Val Gly His Leu Asp  
 35 40 45  
 Leu Leu Glu Ala Ala Lys Ser Ala Leu Asp Arg Gly Gly Tyr Asp Gly  
 50 55 60  
 Leu Phe Leu Gly Gly Asn Tyr Val Ala Gly Val Ala Leu Gly Arg Cys  
 65 70 75 80  
 Val Glu Gly Ala Tyr Glu Ser Ala Ser Gln Ile Tyr Asp Phe Leu Thr  
 85 90 95  
 Lys Tyr Ala Tyr Lys  
 100

005021-52506250